



**Mafalda Filipa Caldeira Camilo**

Licenciada em Engenharia Informática

## **Análise de padrões dos pedidos de intervenção do Regimento de Sapadores Bombeiros**

Dissertação para obtenção do Grau de Mestre em  
**Engenharia Informática**

Orientador: João Carlos Gomes Moura Pires, Professor Associado, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Co-orientadores: Carlos Augusto Isaac Piló Viegas Damásio, Professor Associado, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa  
Fernando Pedro Reino da Silva Birra, Professor Auxiliar, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa

Júri

Presidente: Miguel Carlos Pacheco Afonso Goulão  
Arguente: Ana Paula Pereira Afonso  
Vogal: João Carlos Gomes Moura Pires



FACULDADE DE  
CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE NOVA DE LISBOA

**Fevereiro, 2021**



## **Análise de padrões dos pedidos de intervenção do Regimento de Sapadores Bombeiros**

Copyright © Mafalda Filipa Caldeira Camilo, Faculdade de Ciências e Tecnologia, Universidade NOVA de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.





*Lorem ipsum.*



## AGRADECIMENTOS

Em primeiro lugar, queria agradecer todo o tempo disponibilizado pela parte da Câmara Municipal de Lisboa e pela parte do Regimento de Sapadores Bombeiros (RSB) para reuniões e esclarecimento de dúvidas que foram surgindo no desenvolvimento da dissertação. Da parte da Câmara Municipal de Lisboa queria agradecer em especial ao Sr. João Tremoceiro e da parte do RSB ao Sr. Comandante e 2º Comandante.

Não poderia deixar de agradecer todo o apoio e tempo disponibilizado pelo Prof. João Moura Pires, Prof. Fernando Birra e Prof. Carlos Damásio a este projeto. Também não posso deixar de agradecer ao Prof. Nuno Datia que acompanhou o trabalho.

Por fim, agradeço a todos os meus colegas de mestrado, especialmente, à Ana Sofia Afonso, João Lucas Dias e Sofia Nascimento que partilharam esta experiência comigo que foi naturalmente diferente devido a todo o constrangimento resultante da pandemia COVID-19.



## RESUMO

---

O **Regimento de Sapadores Bombeiros (RSB)** é um corpo municipal de bombeiros que, ao contrário dos bombeiros voluntários, são profissionais. Este corpo é responsável por trabalhos relacionados com o socorro e apoio à população e acções de protecção civil.

O factor tempo é crucial em situações de emergência. Para melhorar a prestação de serviços de emergência aos cidadãos e para otimizar a distribuição de meios é necessário descobrir padrões nos dados das ocorrências deste regimento. Os tipos de padrões abordados podem ser padrões temporais, espaciais e espaço-temporais. Quando os padrões estiverem definidos é necessário encontrar variáveis que possam explicar esses padrões. Por norma, estas variáveis são identificadas através de conjuntos de dados externos que estão correlacionados com os dados originais como potenciais factores explicativos, por exemplo, dados de meteorologia. Com a descoberta de padrões e variáveis explicativas é possível estimar cenários de intervenção. Podendo estimar-se estes cenários tornará possível melhorar a organização e distribuição dos meios de emergência e, consequentemente, será possível diminuir o tempo de resposta das entidades competentes.

Com dados disponibilizados pela Câmara Municipal de Lisboa, relativos aos pedidos de intervenção recebidos por este regimento entre os anos de 2011 e 2018, foram criadas ferramentas interativas de modo a permitir a deteção de padrões temporais, espaciais e espaço-temporais e, posteriormente, proceder à explicação dos mesmos. O objectivo final é disponibilizar ferramentas visuais que contribuam para a deteção e compreensão dos padrões nas ocorrências para melhorar a gestão operacional dos bombeiros. Foram utilizadas técnicas de *clustering* e técnicas de visualização e interação em dados espaço-temporais para o desenvolvimento das ferramentas visuais interativas.

Por fim, as ferramentas desenvolvidas foram avaliadas com um grupo de colegas e professores da faculdade. A avaliação experimental foi realizada remotamente e consistiu num questionário com um conjunto de questões onde as métricas recolhidas para análise foram o tempo de resposta e a taxa de sucesso para cada pergunta.

**Palavras-chave:** Regimento de Sapadores Bombeiros, Análise de Dados, Padrões em Dados, Dados Espaço-temporais, Técnicas de *Clustering*, Visualização Interativa de Dados, Variáveis Explicativas

---

---

## ABSTRACT

---

The Sapper Firefighters Regiment is a municipal fire department which, unlike volunteer firefighters, are professional. This department is responsible for relief and support work to the population and civil protection actions.

The time factor in an emergency is crucial. To improve the provision of emergency services to citizens and to optimize the distribution of means, it is necessary to discover patterns in the occurrence data of this regiment. The types of patterns covered can be temporal, spatial and spatial-temporal. When patterns are defined, it is necessary to find variables that can explain these patterns. Typically, these variables are identified through external data sets that are correlated with the original data as potential explanatory factors, for example, weather data. With the discovery of patterns and the explanatory variables, it is possible to estimate intervention scenarios. Being able to estimate these scenarios will make it possible to improve the organization and distribution of emergency means and, consequently, it will be possible to decrease the response time of the competent entities.

With data provided by the Lisbon City Council, concerning the intervention requests received by this regiment between the years 2011 and 2018, interactive tools were created to allow the detection of temporal, spatial and spatio-temporal patterns and, subsequently, to explain them. The ultimate goal is to provide visual tools that contribute to the detection and understanding of patterns in events to improve the operational management of firefighters. Clustering techniques and visualization and interaction techniques were used in spatio-temporal data for the development of interactive visual tools.

Finally, the tools developed were evaluated with a group of colleagues and professors from the college. The experimental evaluation was carried out remotely and consisted of a questionnaire with a set of questions where the metrics collected for analysis were the response time and the success rate for each question.

**Keywords:** Sapper Firefighter Regiment, Data Analysis, Patterns in Data, Spatiotemporal Data, Clustering Techniques, Data Visualization, Explanatory Variables

---





# ÍNDICE

<b>Lista de Figuras</b>	<b>xvii</b>
<b>Lista de Tabelas</b>	<b>xxiii</b>
<b>Siglas</b>	<b>xxv</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Contexto e Motivação . . . . .	1
1.2 Descrição do Problema . . . . .	2
1.3 Objetivo e Contributos . . . . .	3
1.4 Organização do Documento . . . . .	4
<b>2 Trabalho Relacionado e Estado da Arte</b>	<b>7</b>
2.1 Introdução . . . . .	7
2.2 Trabalhos Relacionados . . . . .	8
2.2.1 Padrões em Incidentes de Incêndios . . . . .	9
2.2.2 Padrões das Chamadas de Emergência . . . . .	10
2.2.3 Padrões de Crimes . . . . .	12
2.2.4 Conclusões . . . . .	13
2.3 Técnicas de <i>Clustering</i> . . . . .	13
2.3.1 Kernel Density Estimation . . . . .	14
2.3.2 Kmeans . . . . .	14
2.3.3 Getis-Ord Gi* do ArcGIS . . . . .	15
2.3.4 Space-time Scan Statistics . . . . .	16
2.3.5 DBSCAN . . . . .	16
2.3.6 SNN . . . . .	17
2.3.7 Conclusões . . . . .	19
2.4 Técnicas de Visualização e Interação de Dados . . . . .	20
2.4.1 Temporal . . . . .	20
2.4.2 Espacial . . . . .	21
2.4.3 Espaço-Temporal . . . . .	22
2.4.4 Interações com visualizações espaço-temporais . . . . .	25
2.4.5 Conclusões . . . . .	28

<b>3</b>	<b>Trabalho Realizado</b>	<b>31</b>
3.1	Abordagem . . . . .	31
3.2	Tecnologias Utilizadas . . . . .	35
3.2.1	Tableau Desktop . . . . .	36
3.2.2	Tableau Online . . . . .	37
3.2.3	Tableau Prep . . . . .	37
3.2.4	QGIS . . . . .	38
3.2.5	Anaconda . . . . .	39
3.2.6	Draw.io . . . . .	39
3.3	Processamento de Dados . . . . .	40
3.3.1	Dados das Ocorrências do RSB . . . . .	40
3.3.2	Dados de Meteorologia . . . . .	46
3.4	Agrupamento . . . . .	48
3.4.1	Temporal . . . . .	48
3.4.2	Espacial . . . . .	52
3.5	<i>Dashboards</i> Desenvolvidos . . . . .	56
3.5.1	Introdução à Linguagem gráfica IVML . . . . .	56
3.5.2	Comparação das Ocorrências nas Áreas de Influência dos Quartéis . . . . .	58
3.5.3	Comparação das Ocorrências nas Freguesias de Lisboa . . . . .	66
3.5.4	Evolução Espacial das Tipologias das Ocorrências ao Longo do Tempo . . . . .	69
3.5.5	Análise do Comportamento Temporal das Tipologias ao Longo de Várias Granularidades Temporais . . . . .	79
3.5.6	Análise do Comportamento Espacial das Ocorrências . . . . .	83
3.5.7	Análise das Ocorrências em Redor de Pontos de Interesse em Lisboa . . . . .	89
3.5.8	Procura de variáveis explicativas para as ocorrências do RSB - Meteorologia . . . . .	95
3.5.9	Conclusões . . . . .	100
3.6	Algumas Análises com as Ferramentas Desenvolvidas . . . . .	102
<b>4</b>	<b>Avaliação do Trabalho Realizado</b>	<b>113</b>
4.1	Avaliação Experimental . . . . .	113
4.2	Participantes . . . . .	118
4.3	Resultados . . . . .	119
4.4	Conclusões . . . . .	128
<b>5</b>	<b>Conclusões e Trabalho Futuro</b>	<b>131</b>
	<b>Bibliografia</b>	<b>133</b>
<b>A</b>	<b>Relatório de qualidade dos dados</b>	<b>139</b>
A.1	Reunir dados iniciais . . . . .	140
A.2	Descrição dos dados . . . . .	140

A.3	Exploração dos dados . . . . .	141
A.3.1	Identificador da ocorrência . . . . .	142
A.3.2	Data de ocorrência . . . . .	142
A.3.3	Latitude e longitude . . . . .	143
A.3.4	Morada . . . . .	144
A.3.5	Freguesia . . . . .	146
A.3.6	Estado da ocorrência . . . . .	147
A.3.7	Tipologia da ocorrência . . . . .	148
A.3.8	Entidades . . . . .	149
A.3.9	Indicador do grau de risco . . . . .	149
A.3.10	Indicador de suspensão antes da activação de meios . . . . .	150
A.3.11	Indicador de falso alarme depois da activação de meios . . . . .	151
A.3.12	Número de viaturas . . . . .	151
A.3.13	Número estimado de elementos . . . . .	152
A.3.14	Descrição da ocorrência . . . . .	154
A.4	Verificação da qualidade dos dados . . . . .	154
A.4.1	Completude . . . . .	154
A.4.2	Consistência . . . . .	156
A.4.3	Validação . . . . .	156
A.4.4	Precisão . . . . .	157
A.4.5	Outras notas . . . . .	158
<b>B</b>	<b>Listagem dos Subtipos das Tipologias Gerais das Ocorrências</b>	<b>159</b>
<b>C</b>	<b>Listagem dos diagramas IVML dos <i>dashboards</i> desenvolvidos</b>	<b>163</b>



## LISTA DE FIGURAS

2.1	Diferença de resultados obtidos entre o algoritmo de <i>clustering Density-based spatial clustering of applications with noise</i> (DBSCAN) e o <i>Shared Nearest Neighbor</i> (SNN) para o mesmo conjunto de dados [15]. . . . .	18
2.2	Exemplo de gráfico de radar. . . . .	20
2.3	Exemplo de mapa de calor. . . . .	21
2.4	Mapa de pontos e mapa de densidade. . . . .	22
2.5	Exemplo de mapa coroplético [12]. . . . .	22
2.6	Estrutura do <i>comap</i> [11]. . . . .	23
2.7	Exemplo do <i>comap</i> [11]. . . . .	24
2.8	Exemplo de uma isosuperfície [6]. . . . .	25
2.9	Exemplo da utilização do cubo espaço-tempo para visualizar grupos [29]. . .	25
3.1	Diagrama de fluxo do processo <i>CRISP-DM</i> [18]. . . . .	32
3.2	Distribuição do número de ocorrências pela tipologia geral. . . . .	43
3.3	Evolução temporal das tipologias gerais das ocorrências ao longo de todas as semanas existentes no conjunto de dados. . . . .	44
3.4	Localização das estações meteorológicas automáticas do Instituto Português do Mar e da Atmosfera (IPMA) em Lisboa. . . . .	47
3.5	Gráfico do método do cotovelo para o conjunto de dados do <i>clustering</i> temporal das tipologias para as estações do ano. . . . .	50
3.6	Resultados do agrupamento temporal. . . . .	51
3.7	Gráfico do método do cotovelo para o conjunto de dados de todas as ocorrências do <i>clustering</i> espacial. . . . .	53
3.8	Resultados do agrupamento espacial. . . . .	55
3.9	Exemplo dos componentes da linguagem <i>Interaction Visualization Modeling Language</i> (IVML). . . . .	57
3.10	Representação de algumas interações. . . . .	58
3.11	Representação dos tipos de ações para iniciar as interações. . . . .	58
3.12	Exemplo de interações iniciadas por diferentes tipos de ações. . . . .	59
3.13	Áreas de influência dos quartéis do RSB. . . . .	59
3.14	<i>Dashboard</i> de comparação das ocorrências nas áreas de influência dos quartéis do RSB. . . . .	60

3.15 Diagrama IVML de interações do <i>dashboard</i> de comparação das ocorrências nas áreas de influência dos quartéis do RSB. . . . .	61
3.16 Exemplo do filtro F1 que contém uma lista com todas as tipologias das ocorrências do RSB e é possível selecionar ou remover tipologias para filtrar os dados. . . . .	63
3.17 Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> de comparação das ocorrências nas áreas de influência dos quartéis do RSB. . . . .	63
3.18 Exemplo de uma <i>tooltip</i> , <i>tooltip</i> T5 do <i>dashboard</i> de comparação das ocorrências nas áreas de influência dos quartéis do RSB. . . . .	64
3.19 Imagem do <i>dashboard</i> de comparação das ocorrências nas áreas de influência dos quartéis do RSB com indicação das visualizações que respondem às questões. . . . .	65
3.20 <i>Dashboard</i> de comparação das ocorrências nas freguesias do concelho de Lisboa. . . . .	66
3.21 Diagrama IVML de interações do <i>dashboard</i> de comparação das ocorrências nas freguesias do concelho de Lisboa. . . . .	67
3.22 Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> de comparação das ocorrências nas freguesias do concelho de Lisboa. . . . .	68
3.23 Imagem do <i>dashboard</i> de comparação das ocorrências nas freguesias de Lisboa com indicação das visualizações que respondem às questões. . . . .	69
3.24 <i>Dashboard</i> da evolução espacial das tipologias das ocorrências ao longo dos anos. . . . .	70
3.25 Diagrama IVML de interações do <i>dashboard</i> da evolução espacial das tipologias das ocorrências ao longo dos anos. . . . .	71
3.26 Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> da evolução espacial das tipologias das ocorrências ao longo dos anos. . . . .	73
3.27 Exemplo da <i>tooltip</i> T1 do <i>dashboard</i> da evolução espacial das tipologias ao longo do tempo. . . . .	74
3.28 <i>Dashboard</i> da evolução espacial das tipologias das ocorrências ao longo dos trimestres. . . . .	74
3.29 Diagrama IVML de interações do <i>dashboard</i> da evolução espacial das tipologias das ocorrências ao longo dos trimestres. . . . .	75
3.30 Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> da evolução espacial das tipologias das ocorrências ao longo dos trimestres. . . . .	76
3.31 <i>Dashboard</i> da evolução espacial das tipologias das ocorrências ao longo dos dias da semana. . . . .	76
3.32 Diagrama IVML de interações do <i>dashboard</i> da evolução espacial das tipologias das ocorrências ao longo dos dias da semana. . . . .	77
3.33 Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> da evolução espacial das tipologias das ocorrências ao longo dos dias da semana. . . . .	78
3.34 Imagem do <i>dashboard</i> de evolução espacial das tipologias das ocorrências ao longo dos anos com indicação das visualizações que respondem às questões. . . . .	79

3.35 Dashboard de análise do comportamento temporal das tipologias ao longo de várias granularidades temporais (estações do ano). . . . .	80
3.36 Diagrama IVML de interações do <i>dashboard</i> de análise do comportamento temporal das tipologias ao longo das estações do ano. . . . .	81
3.37 Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> de análise do comportamento temporal das tipologias ao longo das estações do ano. . . . .	82
3.38 Imagem do <i>dashboard</i> da análise do comportamento temporal das tipologias ao longo das estações do ano com indicação das visualizações que respondem às questões. . . . .	83
3.39 Dashboard de análise do comportamento espacial das ocorrências. . . . .	84
3.40 Diagrama IVML de interações do <i>dashboard</i> da evolução espacial das tipologias das ocorrências ao longo do tempo. . . . .	85
3.41 Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> da evolução espacial das tipologias das ocorrências ao longo do tempo. . . . .	86
3.42 Imagem do <i>dashboard</i> da análise do comportamento espacial das ocorrências do RSB com indicação das visualizações que respondem às questões 1, 2 e 5. . . . .	88
3.43 Imagem do <i>dashboard</i> da análise do comportamento espacial das ocorrências do RSB com indicação das visualizações que respondem à questão 3. . . . .	88
3.44 Imagem do <i>dashboard</i> da análise do comportamento espacial das ocorrências do RSB com indicação das visualizações que respondem à questão 4. . . . .	89
3.45 Dashboard de análise das ocorrências em redor de pontos de interesse no concelho de Lisboa. . . . .	90
3.46 Diagrama IVML de interações do <i>dashboard</i> de análise das ocorrências em redor de pontos de interesse no concelho de Lisboa. . . . .	91
3.47 Dashboard de análise das ocorrências em redor de pontos de interesse no concelho de Lisboa após interação de escolha de um ponto de interesse. . . . .	93
3.48 Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> de análise das ocorrências em redor de pontos de interesse no concelho de Lisboa. . . . .	93
3.49 Imagem do <i>dashboard</i> de análise das ocorrências em redor de pontos de interesse no concelho de Lisboa com indicação das visualizações que respondem às questões. . . . .	95
3.50 Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> de análise das ocorrências em redor de pontos de interesse no concelho de Lisboa. . . . .	96
3.51 Diagrama IVML de interações do <i>dashboard</i> da procura de variáveis explicativas para as ocorrências do RSB - Meteorologia. . . . .	97
3.52 Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> da procura de variáveis explicativas para as ocorrências do RSB - Meteorologia. . . . .	99
3.53 Imagem do <i>dashboard</i> de procura de variáveis explicativas meteorológicas para as ocorrências do RSB com indicação das visualizações que respondem às questões. . . . .	100

3.54	Exemplo de análise do <i>dashboard</i> da evolução espacial das tipologias ao longo dos anos. . . . .	102
3.55	Exemplo de análise do <i>dashboard</i> da evolução espacial das tipologias ao longo dos trimestres. . . . .	103
3.56	Exemplo de análise do <i>dashboard</i> de comparação das ocorrências nas áreas de influência dos quartéis. . . . .	104
3.57	Exemplo de análise do <i>dashboard</i> de análise do comportamento temporal das tipologias do grupo 1 ao longo das estações do ano. . . . .	105
3.58	Exemplo de análise do <i>dashboard</i> de análise do comportamento temporal das tipologias do grupo 2 ao longo das estações do ano. . . . .	106
3.59	Exemplo de análise do <i>dashboard</i> de análise do comportamento dos grupos 0 e 11 resultantes do agrupamento espacial das ocorrências de emergência. . . .	107
3.60	Exemplo de análise do <i>dashboard</i> de análise do comportamento dos grupos 0 e 11 resultantes do agrupamento espacial das ocorrências de emergência ao longo dos dias da semana. . . . .	107
3.61	Exemplo de análise do <i>dashboard</i> de análise do comportamento dos grupos 0 e 11 resultantes do agrupamento espacial das ocorrências de emergência ao longo das horas. . . . .	108
3.62	Exemplo de análise do <i>dashboard</i> de análise das ocorrências em redor de pontos de interesse (transportes). . . . .	109
3.63	Exemplo de análise do <i>dashboard</i> de análise das ocorrências em redor de um ponto de interesse. . . . .	109
3.64	Exemplo de análise do <i>dashboard</i> de procura de variáveis explicativas para as ocorrências do RSB - Meteorologia. . . . .	111
3.65	Exemplo de análise do <i>dashboard</i> de procura de variáveis explicativas para as ocorrências do RSB - Meteorologia - Chuva moderada e chuva forte. . . . .	111
4.1	Fluxo da avaliação experimental. . . . .	117
4.2	Dados demográficos dos participantes. . . . .	118
4.3	Experiência em visualização de dados dos participantes. . . . .	119
4.4	Quadro geral de acertos das perguntas. . . . .	120
4.5	Taxa de sucesso das categorias das perguntas. . . . .	120
4.6	Taxa de sucesso das respostas dos utilizadores com e sem experiência em visualização de dados. . . . .	121
4.7	Média do tempo de resposta por <i>dashboard</i> para respostas corretas e erradas e tempo médio de resposta às perguntas dos <i>dashboards</i> . . . . .	122
4.8	Exemplo do <i>dashboard</i> de agrupamento espacial que necessitou de duas imagens para o questionário. . . . .	123
4.9	Tempo de resposta dos utilizadores por <i>dashboard</i> , identificador imagem e pergunta. . . . .	124
4.10	Tempo de resposta por utilizadores e por género. . . . .	125



4.11	Tempo de resposta por utilizadores e por experiência em visualização de dados.	126
4.12	Média do tempo de resposta por <i>dashboard</i> , pergunta e ordem pela qual aparece no questionário.	127
4.13	Tempo de resposta por utilizadores e pela ordem que aparecem no <i>dashboard</i> das Áreas de Influência.	127
4.14	Visualização com resposta à pergunta P11.	129
4.15	Visualização com resposta à pergunta P06.	129
4.16	Visualização com resposta à pergunta P21.	129
A.1	Evolução temporal, ao longo dos anos, do número de ocorrências.	143
A.2	Número de ocorrências com e sem dados de latitude e longitude.	144
A.3	Distribuição das ocorrências sem latitude e longitude ao longo dos anos.	144
A.4	Número de ocorrências com e sem morada.	145
A.5	Número de ocorrências com e sem código SIG.	145
A.6	Distribuição dos valores de código de fracção.	146
A.7	Número de ocorrências que apresentam e não apresentam identificadores da freguesia.	146
A.8	Número de ocorrências que apresentam e não apresentam designação da freguesia.	146
A.9	Distribuição da descrição do estado da ocorrência.	147
A.10	Número de ocorrências que apresentam e não apresentam descrição do tipo de ocorrência.	148
A.11	Distribuição das ocorrências que não apresentam descrição da sua tipologia ao longo dos anos.	148
A.12	Distribuição dos valores da entidade que deu início à ocorrência.	150
A.13	Distribuição dos valores do grau de risco da ocorrência.	150
A.14	Distribuição dos valores do indicador de suspensão antes da activação de meios.	151
A.15	Distribuição dos valores de falso alarme depois da activação de meios.	151
A.16	Distribuição dos valores do número de viaturas enviadas para o local.	152
A.17	Relação entre nenhuma viatura enviada para o local e a suspensão da ocorrência.	152
A.18	Distribuição dos valores da estimativa do número de elementos enviados para o local.	153
A.19	Relação entre os atributos número estimado de elementos, número de viaturas e tipologia.	153
C.1	Diagrama IVML de interações do <i>dashboard</i> de comparação das ocorrências nas áreas de influência dos quartéis do RSB.	164
C.2	Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> de comparação das ocorrências nas áreas de influência dos quartéis do RSB.	165

C.3	Diagrama IVML de interações do <i>dashboard</i> de comparação das ocorrências nas freguesias do concelho de Lisboa. . . . .	166
C.4	Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> de comparação das ocorrências nas freguesias do concelho de Lisboa. . . . .	167
C.5	Diagrama IVML de interações do <i>dashboard</i> da evolução espacial das tipologias das ocorrências ao longo dos anos. . . . .	168
C.6	Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> da evolução espacial das tipologias das ocorrências ao longo dos anos. . . . .	169
C.7	Diagrama IVML de interações do <i>dashboard</i> da evolução espacial das tipologias das ocorrências ao longo dos trimestres. . . . .	170
C.8	Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> da evolução espacial das tipologias das ocorrências ao longo dos trimestres. . . . .	171
C.9	Diagrama IVML de interações do <i>dashboard</i> da evolução espacial das tipologias das ocorrências ao longo dos dias da semana. . . . .	172
C.10	Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> da evolução espacial das tipologias das ocorrências ao longo dos dias da semana. . . . .	173
C.11	Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> de análise do comportamento temporal das tipologias ao longo das estações do ano. . . . .	174
C.12	Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> de análise do comportamento temporal das tipologias ao longo das estações do ano. . . . .	175
C.13	Diagrama IVML de interações do <i>dashboard</i> da evolução espacial das tipologias das ocorrências ao longo do tempo. . . . .	176
C.14	Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> da evolução espacial das tipologias das ocorrências ao longo do tempo. . . . .	177
C.15	Diagrama IVML de interações do <i>dashboard</i> de análise das ocorrências em redor de pontos de interesse no concelho de Lisboa. . . . .	178
C.16	Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> de análise das ocorrências em redor de pontos de interesse no concelho de Lisboa. . . . .	179
C.17	Diagrama IVML de interações do <i>dashboard</i> da procura de variáveis explicativas para as ocorrências do RSB - Meteorologia. . . . .	180
C.18	Diagrama IVML de <i>tooltips</i> do <i>dashboard</i> da procura de variáveis explicativas para as ocorrências do RSB - Meteorologia. . . . .	181

## LISTA DE TABELAS

2.1	Taxonomia de interação de visualização proposta por Figueiras [33]. . . . .	26
2.2	Taxonomia de interação de visualizações espaço-temporais [33]. . . . .	29
3.1	Informação de localização das ocorrências . . . . .	40
3.2	Informação temporal das ocorrências. . . . .	41
3.3	Informação das características das ocorrências. . . . .	42
3.4	Tipologias gerais presentes no conjunto de dados das ocorrências. . . . .	43
3.5	Identificadores de tipologias a mapear para diferentes descrições de tipologias. . . . .	46
3.6	Atributos presentes no conjunto de dados meteorológicos do IPMA. . . . .	46
3.7	Colunas do novo conjunto de dados após processamento. . . . .	48
3.8	Decomposição de um dia por partes. . . . .	49
3.9	Exemplo do conjunto de dados criado para o <i>clustering</i> das tipologias para as estações do ano. . . . .	49
3.10	Valores da métrica <i>silhouette score</i> para diferentes $k$ no agrupamento por estações do ano. . . . .	50
3.11	Síntese do agrupamento temporal. . . . .	51
3.12	Tipologias das ocorrências consideradas de emergência. . . . .	53
3.13	Valores da métrica <i>silhouette score</i> para diferentes <i>minPoints</i> no agrupamento com dados das ocorrências de emergência. . . . .	54
3.14	Síntese do agrupamento espacial. . . . .	54
3.15	Atributos do conjunto de dados dos pontos de interesse. . . . .	90
3.16	Caracterização dos valores relativos à humidade. . . . .	95
3.17	Caracterização dos valores relativos à precipitação. . . . .	96
3.18	Caracterização dos valores relativos à temperatura. . . . .	96
3.19	Caracterização dos valores relativos à intensidade do vento. . . . .	96
4.1	Organização das perguntas de acordo com as ferramentas de deteção de padrões a serem avaliadas, as visualizações alvo e as questões aos utilizadores. . . . .	115
4.2	Organização das perguntas de acordo com as ferramentas de deteção de variáveis explicativas a serem avaliadas, as visualizações alvo e as questões aos utilizadores. . . . .	116
A.1	Significado das colunas dos dados . . . . .	141

## LISTA DE TABELAS

---

A.2	Tipos de valor dos atributos . . . . .	142
A.3	Subtipos das ocorrências da tipologia geral "Pré-Hospitalar" . . . . .	149
B.1	Subtipos das diversas tipologias gerais presentes no conjunto de dados das ocorrências do RSB. . . . .	162

## SIGLAS

CML	Câmara Municipal de Lisboa.
CODU	Centros de Orientação de Doentes Urgentes.
DBSCAN	<i>Density-based spatial clustering of applications with noise.</i>
INEM	Instituto Nacional de Emergência Médica.
IPMA	Instituto Português do Mar e da Atmosfera.
IVML	<i>Interaction Visualization Modeling Language.</i>
KDE	<i>Kernel Density Estimation.</i>
RSB	Regimento de Sapadores Bombeiros.
SNN	<i>Shared Nearest Neighbor.</i>
ST-SNN	<i>Spatial-temporal Shared Nearest Neighbor.</i>
ST-DBSCAN	<i>Spatial-temporal Density-based spatial clustering of applications with noise.</i>
STKDE	<i>Spatial-temporal Kernel Density Estimation.</i>



## INTRODUÇÃO

O Regimento de Sapadores Bombeiros (RSB) é o corpo municipal de bombeiros de Lisboa. Estes bombeiros são profissionais e existem várias equipas de resposta de emergência especializadas. As principais missões são de socorro à população e a náufragos e o combate a incêndios. Também fazem apoio a acções de protecção civil. Estas missões têm como objectivo assegurar as actividades de protecção e socorro em Lisboa. Existem vários órgãos de execução, uma companhia de Comando e Serviços, outra companhia de Intervenção Especial, uma escola de Sapadores Bombeiros e actualmente existem sete quartéis das companhias de Sapadores Bombeiros em Lisboa [25].

O RSB recebe diversos tipos de pedidos de intervenção. Estes pedidos podem ser o clássico resgate de animais, mas podem também ser emergências médicas ou incêndios. Este regimento atende a pedidos de particulares que solicitam a sua presença e colaboração atempadamente. Este é o caso de acções de prevenção em eventos, como por exemplo, jogos de futebol [25]. Estes pedidos diferem de outro tipos, pois estão programados.

### 1.1 Contexto e Motivação

Esta dissertação resulta de uma parceria entre a [Câmara Municipal de Lisboa \(CML\)](#) e a Faculdade de Ciências e Tecnologias da Universidade Nova de Lisboa (FCT NOVA), no âmbito do projecto Laboratório de Dados Urbanos de Lisboa (*LxDataLab*)<sup>1</sup>. O *LxDataLab* apresenta vários desafios sobre diferentes problemáticas da cidade de Lisboa. Estes desafios têm como objectivo a criação de várias soluções, nomeadamente, de analítica e visualização de grandes volumes de dados que permitam melhorar o planeamento e gestão operacional da cidade [27]. O tema desta dissertação é um dos desafios propostos pelo *LxDataLab*.

---

<sup>1</sup><https://lisboainteligente.cm-lisboa.pt/lxi-iniciativas/lx-data-lab/>

A principal motivação deste desafio é ajudar a compreender a distribuição temporal, espacial e temática dos vários tipos de pedidos de intervenção do Regimento de Sapadores Bombeiros e compreender se essas distribuições se alteram sob determinadas circunstâncias, como por exemplo, eventos meteorológicos. Esta compreensão ajudará os bombeiros a entender as necessidades da cidade de Lisboa sob determinadas circunstâncias. Sabendo quais são as áreas e horas com mais incidência dos vários tipos de ocorrências, os bombeiros conseguirão melhorar a sua gestão operacional, isto é, conseguirão melhorar a alocação dos vários tipos de meios de emergência e, consequentemente, o tempo de resposta a essas intervenções diminuirá. Sendo o tempo um factor crucial num cenário de emergência a sua diminuição tem impacto na vida dos cidadãos da cidade. A diminuição do tempo de resposta às intervenções do Regimento de Sapadores Bombeiros resulta numa melhoria da prestação de serviços de emergência aos cidadãos.

### 1.2 Descrição do Problema

Em situações de emergência, o tempo de resposta das entidades de socorro é crucial. O tempo de resposta às intervenções é, em grande parte, influenciado pela distribuição dos meios de emergência pela cidade. Uma melhor organização e distribuição dos meios de socorro permite que o tempo de resposta seja menor. Esta organização pode ser melhorada através da detecção e compreensão de padrões nos pedidos de intervenção. Desta forma os meios chegarão mais rapidamente ao local da intervenção e os cidadãos receberão um melhor serviço de emergência.

Os dados disponibilizados, para a realização desta dissertação foram cedidos pela Câmara Municipal de Lisboa (CML). Estes dados são referentes aos pedidos de intervenção recebidos pelo Regimento de Sapadores Bombeiros (RSB) entre os anos de 2011 e 2018.

É pretendido, com os dados das ocorrências do RSB, desenvolver ferramentas de análise de modo a detectar padrões nas intervenções deste regimento. Padrões em dados definem a forma pela qual os dados podem ser estruturados e indexados, revelando tendências. A ausência de padrões nos dados demonstra comportamentos aleatórios. Os padrões que poderão ser detectados neste conjunto de dados são padrões temporais, espaciais ou uma conjunção destes.

Com as ferramentas desenvolvidas e com os possíveis padrões detectados através delas, será feita uma análise para encontrar factores explicativos para esses padrões. Uma variável explicativa, por norma, é identificada através de conjuntos de dados externos aos dados das ocorrências, por exemplo, dados de condições meteorológicas. Estas variáveis estarão correlacionadas, ou fortemente correlacionadas, como potenciais factores explicativos dos padrões.

Com a detecção dos padrões e de variáveis que ajudem a explicar os mesmos será possível estimar cenários de intervenção. O RSB, com a compreensão e explicação dos padrões detectados nos dados das ocorrências, pode estimar cenários de intervenção. Isto possibilita uma melhoria na organização dos seus meios de emergência pela cidade de



Lisboa. Estando os meios posicionados em locais onde podem acontecer ocorrências com certas características, sob certas condições, o tempo de resposta da entidade competente poderá ser reduzido e, assim, serão melhorados os serviços de emergência prestados aos cidadãos.

É importante salientar que nesta dissertação não existe um panorama completo de todas as respostas às emergências da cidade de Lisboa. Isto porque existem outras forças de emergência como, por exemplo, o [Instituto Nacional de Emergência Médica \(INEM\)](#) que não serão abordadas. Para a realização desta dissertação apenas vão ser usados dados referentes às intervenções feitas pelo Regimento de Sapadores Bombeiros. Outro aspecto importante de salientar, é que o objectivo desta dissertação é, somente, desenvolver ferramentas que permitam a detecção de possíveis padrões existentes nos dados das intervenções. O foco não está na optimização da alocação dos meios de emergência. No entanto, os resultados obtidos nesta dissertação poderão ajudar o [RSB](#) a melhorar a sua organização operacional.

### 1.3 Objetivo e Contributos

O objectivo desta dissertação, de um modo geral, é o desenvolvimento de ferramentas que permitam conhecer e analisar os dados das ocorrências do [RSB](#). Para tal, é necessário fazer um levantamento de um conjunto de questões relevantes para o problema. As ferramentas desenvolvidas serão *dashboards* que serão constituídos por diferentes visualizações que respondam às questões que foram levantadas e assim os bombeiros do [RSB](#) terão mais conhecimentos, nomeadamente, de padrões nos dados e dessa forma poderão tomar decisões mais informadas sobre diversos tópicos como a alocação de meios atempadamente.

Relativamente aos contributos resultantes da realização desta dissertação, as contribuições são as seguintes:

- **Relatório de qualidade de dados**

Esta dissertação é baseada no uso de dados e, como tal, é necessário a familiarização, exploração e análise dos mesmos. Um factor bastante importante, quando se trabalha com dados, é a qualidade destes. Um dos resultados da dissertação foi a elaboração de um relatório de qualidade de dados para análise da qualidade dos dados das ocorrências e para apoio à tomada de decisões em futuras análises. Este relatório beneficiou a [CML](#).

- **Dados processados**

Baseado nos resultados obtidos através da realização do relatório de qualidade dos dados das ocorrências do [RSB](#) foi feito um pré-processamento dos mesmos dados para aumentar a sua qualidade. Este pré-processamento garantiu que os dados que irão ser a base da dissertação têm qualidade suficiente para serem utilizados em

diversas análises. Este processamento de dados ajudará em posteriores estudos ou trabalhos que possam vir a ser desenvolvidos com os mesmos dados.

- **Ferramentas para análise, exploração e detecção de padrões nos dados das ocorrências**

O objectivo da dissertação é o desenvolvimento de ferramentas que permitam a análise, a exploração e a detecção de padrões espaciais, temporais e espaço-temporais. Portanto, foi desenvolvido um conjunto de diferentes *dashboards* que permitem responder a diferentes grupos de perguntas que são relevantes para o problema em questão. Com estes *dashboards* o comando operacional e estratégico do [RSB](#) poderá conhecer e explorar os seus dados das ocorrências e detetar potenciais padrões, podendo assim, tomar decisões de alocação de meios mais informadas.

- **Ferramenta para análise e procura de correlações com dados meteorológico**

Outro objetivo da dissertação é o desenvolvimento de ferramentas para detetar a existência, ou não, de variáveis potencialmente explicativas. Foi desenvolvido um *dashboard* que permite relacionar as ocorrências do [RSB](#) com dados meteorológicos. Com este *dashboard* o comandante do [RSB](#) poderá antecipar cenários de intervenção quando existem certas condições meteorológicas.

- **Linguagem gráfica para auxiliar o processo de *design* e documentação de *dashboards***

Durante o desenvolvimento desta dissertação, devido à falta de metodologias e ferramentas para auxiliar o processo de *design* e de documentação de *dashboards*, conjuntamente com outros colegas de mestrado, foi iniciado o processo de criação de uma nova linguagem gráfica que tem como objetivo combater essa necessidade. Esta linguagem denomina-se [IVML](#) - *Interaction Visualization Modeling Language* e encontra-se atualmente ainda em desenvolvimento.

## 1.4 Organização do Documento

Este documento está estruturado em cinco capítulos, sendo que um deles é o capítulo presente. Os restantes capítulos estão organizados da seguinte forma:

- **Trabalho Relacionado e Estado da arte**

No Capítulo [2](#) é feito um levantamento do estado da arte relativo a este problema. Este capítulo apresenta trabalhos relacionados com a detecção de padrões espaciais e/ou temporais, técnicas de *clustering* e técnicas de visualização e interação de dados.

- **Trabalho Realizado**

No capítulo 3 é apresentado o trabalho que foi desenvolvido nesta dissertação, nomeadamente, qual foi a abordagem seguida para resolver o problema, a apresentação das tecnologias utilizadas para desenvolver as ferramentas criadas, como foi feito o processamento dos conjuntos de dados utilizados e, por fim, é feita uma exposição detalhada das ferramentas (*dashboards*) que foram desenvolvidas.

- **Avaliação do Trabalho Realizado**

O capítulo 4 contém a descrição do processo de avaliação do trabalho que foi desenvolvido durante a dissertação, nomeadamente, das ferramentas criadas. Estas ferramentas foram submetidas para avaliação experimental com utilizadores não experientes na temática.

- **Conclusões e Trabalho Futuro**

No capítulo 5 são apresentadas as conclusões retiradas sobre o trabalho desenvolvido durante a dissertação. Neste capítulo também é feita uma breve descrição do que poderá ser o trabalho futuro sustentado no trabalho que foi desenvolvido.



## TRABALHO RELACIONADO E ESTADO DA ARTE

Neste capítulo será feito o levantamento do estado da arte e serão apresentados os trabalhos relacionados com o tema desta dissertação.

O capítulo tem uma introdução que apresenta como a pesquisa foi efectuada, nomeadamente, quais as fontes utilizadas, palavras-chave e critérios de inclusão e exclusão. De seguida, serão apresentados os trabalhos relacionados, as técnicas de *clustering*, visualização e interação utilizadas por esses trabalhos e consideradas relevantes. Sempre que uma subsecção acabe, no final existem as conclusões que foram obtidas com as pesquisas e leituras efetuadas.

### 2.1 Introdução

O estado da arte permite conhecer trabalhos já elaborados dentro do mesmo tema da dissertação. Também permite ficar a conhecer quais as metodologias desenvolvidas e utilizadas em trabalhos relacionados para resolver problemas semelhantes ou relacionados.

Para a pesquisa da literatura foram utilizadas as seguintes fontes:

- Google Académico;
- ScienceDirect;
- IEEEExplore.

As palavras-chave utilizadas para efectuar as pesquisas da literatura nestas fontes foram: "*patterns in firefighters calls*", "*spatiotemporal patterns*", "*spatiotemporal analysis*", "*spatiotemporal clustering*" e "*spatiotemporal visualization*". Com estas palavras-chave surgiram muitos resultados, mas poucos relacionados com o tema de chamadas de emergências relativas a bombeiros.

Como não foram encontrados muitos trabalhos relacionados com o tema desta dissertação, houve necessidade de criar alguns critérios de inclusão. Foram incluídos outros trabalhos que apresentassem tópicos semelhantes, por exemplo a detecção de padrões em outros tipos de eventos que não chamadas de pedidos de intervenção dos bombeiros. Os eventos considerados foram: chamadas de pedidos de socorro do 112 (ou número equivalente noutros países), ocorrências de crimes e tipos de fogos.

Os critérios de exclusão, dos resultados obtidos nas pesquisas, utilizados foram a data da publicação dos documentos e o resumo. A data foi um dos critérios de exclusão utilizados pois, por exemplo, a área de *machine learning* tem evoluído muito rapidamente e existem novos métodos que são mais poderosos que propostas anteriores. A área da visualização também tem evoluído ultimamente e existem agora novos métodos que são mais poderosos. O resumo foi um dos critérios de exclusão utilizados, pois, por vezes, só o título não chega para entender o conteúdo do trabalho. O número de citações foi um critério, não de exclusão, mas que foi tido em conta.

Este capítulo está dividido por três tópicos principais que são os seguintes:

- **Trabalhos relacionados**

Nesta secção serão apresentados os vários trabalhos relacionados encontrados na literatura e as conclusões dos mesmos.

- **Técnicas de *clustering***

Nesta secção serão apresentados os métodos/algoritmos de *clustering* que foram utilizados nos trabalhos relacionados. Também serão apresentados outros algoritmos que estão relacionados com *clustering* espaço-temporal.

- **Técnicas de visualização e interação**

Nesta secção serão introduzidas visualizações utilizadas na literatura para alcançar os seus objectivos. Também serão apresentadas outras visualizações, que embora não tenham sido referidas nos trabalhos relacionados, permitem visualizar as componentes espaciais e/ou temporais, sendo por isso consideradas relevantes. Por fim, serão levantadas quais as interações com visualizações existentes na literatura relativamente a dados espaço-temporais.

## 2.2 Trabalhos Relacionados

Após várias pesquisas não foram encontrados trabalhos que se relacionassem com a detecção de padrões em pedidos de intervenção de bombeiros. Portanto, houve a necessidade de procurar outros trabalhos que estivessem relacionados com a detecção de padrões, mas noutros conjuntos de dados. Os trabalhos relacionados encontrados assentam em dados que também fazem parte de algumas tipologias das ocorrências que os bombeiros podem responder, por exemplo, incêndios, emergências médicas (chamadas do 112, ou número equivalente) e crimes.

### 2.2.1 Padrões em Incidentes de Incêndios

Existe alguma literatura na detecção de padrões em incidentes de incêndios com o intuito de apoiar a organização dos meios de emergência e socorro. No entanto, a literatura que aborda a detecção destes padrões não está tão desenvolvida como a de outros fenómenos urbanos [12].

Com o intuito de preencher a lacuna existente na literatura, o trabalho [12] levou a cabo uma análise espacial de dados de incêndios domésticos, incêndios em veículos, incêndios secundários e falsos alarmes maliciosos na área sul do País de Gales. A distribuição geográfica destes incidentes sugeriu uma relação com uma variável socioeconómica. Para investigar a relação dos tipos de incidentes e variáveis socioeconómicas foi utilizado o índice de privação de *Townsend*. Este índice tem em conta variáveis de desemprego, ausência de veículo próprio, ausência de casa própria e sobrelotação da habitação [12]. Para cada bairro da área de estudo, foi calculada esta pontuação usando dados do Censos. A análise deste estudo sugere que bairros menos carenciados tendem a ser associados às categorias de menor risco de incidentes e todos os tipos de incidentes exibem uma alta concentração espacial onde existe uma maior concentração da população. Tal como relatado noutro trabalho [22], em dados da cidade de Nova Iorque, existe uma relação entre factores socioeconómicos e risco de incidentes relacionados com incêndios, onde os bairros com mais população e carenciados são mais propensos a conter estes incidentes.

Como o trabalho acima indicado apenas se foca na componente espacial dos dados, o mesmo autor realizou outro trabalho, com os mesmos dados, mas numa perspectiva espaço-temporal [11]. Para a análise temporal foi feita uma assumpção importante: o tempo do registo da chamada a relatar um incidente de incêndio é considerado perto do tempo real da ignição do fogo. Para investigar padrões temporais foram usadas visualizações para diferentes granularidades de tempo (hora, dia e mês). O trabalho [11] mostra que a densidade de certos tipos de incidentes se vai modificando ao longo do tempo, sendo que, existem diferentes áreas de grande densidade entre horas e dias da semana, nomeadamente, dias úteis e fins de semana.

Outro trabalho de análise de padrões espaço-temporais realizado com dados do sudeste de *Queensland*, Austrália, revelou a relação entre certos tipos de incidentes de incêndios, nomeadamente incêndios em propriedades, e o estado sócioecómico [8]. Esta descoberta vai ao encontro de outros estudos [12, 22]. A análise da dinâmica espaço-temporal de incêndios secundários (vegetação e outros incêndios externos) varia de acordo com o mês do ano, o que indica uma possível relação com as características do verão. O estudo em [8] afirma que é necessário desenvolver mais trabalhos onde se introduzam outras variáveis (meteorologia ou feriados) para encontrar explicações para os padrões detectados.

Em Toronto, Canadá, outro trabalho semelhante surgiu e detectou bastantes padrões temporais em incidentes de incêndios de vários tipos e procurou explicações para os mesmos [4]. Segundo o trabalho em [4], incêndios estruturais tornam-se mais frequentes de noite e durante os fins de semana. Possíveis razões, apontadas em [4], incluem a presença

de crianças nas casas e vigilância inadequada em locais de trabalho abandonados. Actividades de vandalismo e incêndio criminoso aumentam também durante a noite, pois existe vantagem de operar sob a cobertura de escuridão, são mais comuns nos meses de verão e são mais frequentes nos fins de semana. Os incêndios causados por crianças ocorrem com maior frequência quando as crianças estão em casa e estão acordadas, das 15 às 20 horas, são mais frequentes durante os dias de semana devido às crianças estarem sozinhas em casa. Os incêndios causados por falhas eléctricas e mau uso ocorrem principalmente durante o final da tarde e o início da noite. As actividades de culinária são os principais contribuintes para incêndios em residências. A análise espacial permitiu concluir que a menor intensidade de incidentes é encontrada em áreas que representam bairros com baixa densidade populacional, construções mais recentes e rendas mais altas. A maior parte dos tipos de incidentes exibe uma alta concentração espacial comum no centro da cidade devido à maior concentração de estruturas envelhecidas, população e actividades. A análise espaço-temporal deste trabalho demonstra que existem alterações dos focos de densidade dos incidentes ao longo do tempo tal como em [11].

Com o intuito de explorar a relação entre incêndios e roubos residenciais foram analisados, espacial e temporalmente, dados da cidade *Surrey*, Canadá [40]. Estes dados foram agregados geograficamente aos limites da área do Censos. A agregação de dados generaliza padrões espaciais, impedindo a interpretação a nível individual e deixando a análise de dados susceptível ao problema de unidade de área modificável, isto é, dados de pontos agregados a um nível espacial diferente podem exibir padrões diferentes [40]. A análise temporal, do trabalho [40], demonstrou que o número de incêndios aumentara entre 13 e as 20 horas e são uniformemente distribuídos ao longo do dia da semana. Em comparação, o roubo aumenta entre as 8 e 19 horas e é menos frequente nos finais de semana. Foram aplicadas estatísticas circulares aos padrões temporais. As estatísticas circulares fornecem uma maneira de analisar dados que circulam sobre si mesmos e podem ser usadas para examinar a força dos padrões temporais. O teste de *Rayleigh* serve para indicar se uma distribuição circular é aleatória ou não [40]. Em incêndios, o teste de *Rayleigh* indicou agrupamentos significativos para a hora, mas não para o dia da semana ou para o mês. Em roubos, este teste indicou um agrupamento significativo na hora, dia da semana e mês. O teste  $U^2$  de *Watson* é um teste não paramétrico que serve para indicar se as distribuições de dois conjuntos de dados circulares são significativamente diferentes [40]. Este teste indicou diferenças significativas nas distribuições ao longo da hora, dia da semana e mês dos incêndios e roubos. Portanto, zonas afectadas por incêndios são também afectadas por roubos, mas os padrões temporais destes tipos de incidentes são distintos.

### 2.2.2 Padrões das Chamadas de Emergência

O tempo entre a chamada do pedido de intervenção e a acção operacional é crucial. A rapidez da resposta dada à chamada depende da disponibilidade de recursos e infraestrutura na área de onde a chamada se originou.



A análise de padrões temporais e/ou espaciais tenta determinar quais os pontos críticos das chamadas e os factores que as influenciam. Estas análises podem ajudar a moldar a elaboração de estratégias, ser um guia para distribuir os recursos de incêndios, médicos e policiais para as comunidades de maneira justa e, potencialmente, salvar vidas, diminuindo o tempo de resposta para as chamadas [35].

Existem trabalhos relacionados com detecção de padrões temporais das chamadas de emergência. O trabalho em [7] faz essa análise em chamadas de emergência médica numa área metropolitana da Austrália. O trabalho [7] conclui que os padrões de distribuição da demanda às sextas, sábados e domingos são significativamente diferentes dos de segunda a quinta-feira. Essas diferenças estão associadas a um aumento do número de casos de trauma por volta da meia-noite às sextas e sábados, sugerindo uma associação entre trauma, idade mais jovem e actividade recreativa. Os casos médicos são mais frequentes durante o horário de trabalho durante a semana e têm uma média de idade mais avançada. Este trabalho também alerta para o cuidado a ter na agregação de dados e na detecção de padrões, tal como em [40].

Existe alguma literatura na detecção de *hotspots* de chamadas de emergência. O trabalho em [21] propõe um algoritmo espaço-temporal para a detecção desses *hotspots*. Os *hotspots* são definidos como um grande número de chamadas que ocorrem a uma certa distância e num determinado período de tempo. Quaisquer duas chamadas que ocorrem dentro da distância mínima e dentro do tempo mínimo uma da outra serão agrupados. Depois de todos os grupos estarem definidos, somente aqueles de um determinado tamanho são seleccionados como *hotspots*. Um *hotspot* irá conter um conjunto de chamadas relacionadas com o mesmo evento. No entanto, pode acontecer que as chamadas individuais não se relacionam entre si, mas simplesmente acontecem juntas no tempo e no espaço em grandes números. A configuração óptima dos parâmetros do algoritmo tenta evitar essas situações. Este algoritmo é capaz de agrupar chamadas que estão, potencialmente, correlacionadas. O trabalho em [21] determinou que a maioria das chamadas, nas cidades de São Francisco e São Diego, Estados Unidos, ocorre um minuto após a anterior e o tempo médio entre chamadas aumenta nas primeiras horas da manhã e é mais baixo durante a tarde.

Compreender os factores que contribuem para altos volumes de chamadas de emergência, ou seja, *hotspots* é importante, pois os resultados podem ajudar a moldar a formulação de estratégias e ser um guia para distribuir recursos para as comunidades de maneira justa. Recomendações para a realocação das forças de serviço de emergência podem ser fornecidas com os resultados da análise de *hotspots* [35].

Por fim, neste campo existem alguns trabalhos que apresentam modelos preditivos, porém estes modelos saem do escopo desta dissertação. No entanto, o trabalho [9], para além de propor um modelo preditivo, faz uma análise temporal de chamadas de emergência e detecção de eventos anómalos. Os padrões temporais destacam os horários de pico da demanda de chamadas de emergência, além de destacarem a diferença de comportamento entre as diferentes zonas da cidade. Os padrões temporais podem ser representados por

séries temporais horárias do total de chamadas de emergência com séries temporais separadas para cada zona da cidade, tipo de chamada e dia da semana [9]. Para o uso desta série temporal, os dados necessitam de ser agregados para criar um perfil acumulativo de 24 horas para cada tipo de chamada, em cada região geográfica, para cada dia da semana. As representações visuais dos padrões das chamadas permitirão que as entidades identifiquem quais as zonas que necessitam de recursos durante os vários turnos, informando as decisões de alocação. Um dos aspectos críticos da alocação eficaz de recursos é saber identificar regiões de interesse que podem exigir uma modificação dos recursos.

### 2.2.3 Padrões de Crimes

Os bombeiros podem responder a ocorrências relacionadas com crimes, por exemplo, agressões e casos de tentativas de homicídio.

A literatura que aborda a detecção de padrões relativos a fenómenos de criminalidade está muito desenvolvida [12]. Existem vários trabalhos sobre o famoso conjunto de dados de crimes na cidade de Chicago. Contudo, menos atenção tem sido dada ao desenvolvimento de análises e métodos representacionais de dimensões temporais, em comparação com as dimensões geográficas de epidemiologia do crime [29].

O trabalho presente em [29] faz uso de dados de ocorrências de delitos relacionados com furto e roubo relatados à polícia de Quioto. O trabalho em [29] fez uma análise espaço-temporal destes eventos que revelou aglomerados estáveis, durante o período de estudo, na parte central da cidade e ao redor da estação de Quioto, bem como aglomerados transitórios em torno de várias estações ferroviárias suburbanas. Diferenças temporais nos grupos transitórios, que alternam entre duas regiões da cidade, sugerem um novo tipo de fenómeno de deslocamento que pode ser causado ou pelo comportamento racional dos criminosos para cometer crimes de forma eficiente ou por acções locais de prevenção ao crime na parte norte da cidade que podem ter empurrado os grupos criminosos para a parte sul da cidade [29]. O trabalho em [29] realça a eficácia da combinação de estatísticas descritivas e confirmativas num ambiente interactivo de visualização para a análise de dados espaço-temporais.

Existem vantagens numa abordagem interdisciplinar, ou seja, trabalhar em conjunto com especialistas no domínio. O trabalho em [30] demonstra a importância do contacto com especialistas, neste caso, de justiça criminal. Esta abordagem permitiu a detecção de um campo de texto livre no conjunto de dados que apresenta informações relevantes para a detecção de padrões. A existência deste campo é desafiante no sentido de converter a informação relevante presente nele em atributos. Esta abordagem também permitiu identificar quais os atributos que são significativos para a formação de grupos de crimes. Este processo envolveu falar com especialistas do domínio e executar iterativamente o algoritmo de importância do atributo para chegar ao conjunto final de atributos para agrupar os tipos de crimes. A técnica usada em [30] foi capaz de identificar os padrões de criminalidade de um grande número de crimes facilitando o trabalho dos detectives.

Alguns trabalhos utilizam outros conjuntos de dados para além dos dados de crimes. O trabalho em [28] utiliza dados de redes sociais, dados de população ambiente dinâmica, para estimar a população ambiente e identificar grupos de crimes temporais no espaço que são significativos, dado o número de potenciais vítimas presentes no momento da ofensa. A localização de grupos vulneráveis de pessoas é altamente dinâmica. Portanto, uma estimativa precisa da verdadeira população em risco num determinado local e tempo é vital para o cálculo confiável da taxa de criminalidade e a geração de *hotspots*. Estes *hotspots* permitirão fazer uma análise espaço-temporal revelando possíveis padrões.

#### 2.2.4 Conclusões

Os resultados dos trabalhos que têm como missão detectar padrões temporais, espaciais e espaço-temporais podem ajudar a moldar a formulação de estratégias e ser um guia para a distribuição justa de recursos às comunidades.

Existem vários tipos de padrões: os temporais; os espaciais e os espaço-temporais. Os padrões espaço-temporais descrevem um comportamento temporal e espacial geral de um grupo de eventos.

Existem várias técnicas para detectar estes padrões. Estas técnicas podem ser técnicas de *clustering* e/ou de visualização.

Relativamente a técnicas de *clustering* para detectar padrões, grande parte da literatura faz uso do *Kernel Density Estimation* (KDE) para *clustering* espacial e existe uma variante espaço-temporal denominado *Spatial-temporal Kernel Density Estimation* (STKDE). O *K-means* também foi utilizado tanto para agrupar dados geograficamente como para séries temporais. Para a detecção de *hotspots* a estatística *Getis-Ord Gi\** foi utilizada. Por fim, a estatística *Space-Time Scan Statistic* também foi usada em alguns trabalhos relacionados.

As técnicas de visualização podem ser utilizadas para detectar padrões temporais e/ou espaciais. O uso de mapas de densidade, com o resultado do algoritmo de *clustering* KDE, é um dos principais métodos de visualização espacial. Outra visualização muito utilizada para detectar padrões espaciais é o mapa coroplético. Para a análise da componente temporal os gráficos mais usados são o gráfico de linhas e o gráfico de radar. O uso de *comaps* para visualizar a componente espacial e temporal, simultaneamente, também tem sido muito popular. Em trabalhos mais recentes, o uso do cubo espaço-tempo, juntamente com as técnicas STKDE e *Space-Time Scan Statistic*, tem sido aplicado a dados espaço-temporais para a procura de padrões também espaço-temporais.

### 2.3 Técnicas de Clustering

Nesta secção serão apresentados os métodos e algoritmos de *clustering* que foram utilizados nos trabalhos relacionados para a detecção de padrões. Também serão apresentados outros métodos que, apesar de não terem sido utilizados nos trabalhos relacionados, existem e podem ser aplicados para agrupar dados espaço-temporais. O agrupamento

espácio-temporal está a ganhar mais atenção científica e tem como objectivo agrupar objectos com base na sua semelhança espacial e temporal, ajudando a descobrir padrões e correlações temporais e espaciais.

### 2.3.1 Kernel Density Estimation

O KDE é amplamente utilizado para produzir uma superfície suavizada numa série de pontos individuais. Esta estimativa é não paramétrica e é utilizada quando se pretende estimar uma função de densidade. Existem várias funções do *kernel* como, por exemplo, a função de *Gauss* ou a função uniforme. No entanto, o parâmetro mais influente no KDE é a escolha da largura de banda. O parâmetro da largura de banda controla a suavidade da estimativa de densidade. Valores da largura de banda muito baixos irão produzir estimativas de densidade pontiagudas, onde cada pico está associado a um ponto de dados individual, enquanto valores altos de largura de banda irão produzir estimativas de densidade mais suaves [12]. Uma largura de banda aceitável é aquela em que a saída resultante não é muito pontiaguda nem muito suave.

O KDE é muito utilizado para a produção de mapas de densidade [8, 11, 12], onde é possível identificar áreas de maior e menor densidade de acontecimentos.

Mais recentemente, foi proposta uma extensão do KDE, o STKDE [6]. Esta proposta considera cada recurso como um ponto no espaço-tempo  $(x, y, t)$ . São utilizadas três dimensões, a componente espacial é geralmente considerada bidimensional e a componente temporal será a dimensão restante. Este triplo é então usado como entrada para o algoritmo tridimensional da estimativa de densidade *kernel* que tenta estimar a função de densidade de probabilidade  $f(x, y, t)$ . O valor dessa função reflecte a probabilidade de um evento ocorrer no local  $(x, y)$  no momento  $t$  [6].

### 2.3.2 Kmeans

O *Kmeans* é um dos algoritmos de partição mais famoso e simples [20]. Os algoritmos de *clustering* particional localizam todos os grupos simultaneamente como uma partição dos dados.

Este algoritmo de *clustering* consiste em dividir o conjunto de dados em  $k$  grupos, cada um definido pelo vector médio dos membros do grupo, que é o protótipo desse grupo. Cada exemplo pertence ao grupo representado pelo protótipo mais próximo. Sendo assim, o *Kmeans* é considerado um método de agrupamento exclusivo, particional e baseado em protótipos. O algoritmo para calcular o grupo de *Kmeans* começa com um conjunto aleatório de  $k$  protótipos, depois cada exemplo é atribuído ao protótipo mais próximo, de seguida é necessário computar de novo cada protótipo como o ponto médio de todos os exemplos atribuídos a esse grupo. O algoritmo continua a atribuir protótipos aos exemplos e a computar os novos protótipos até convergir ou atingir algum critério de paragem [20].

O *Kmeans* tem três parâmetros: o valor de  $k$ , o método de inicialização e a medida de distância. A escolha mais crítica é a do valor de  $k$  que representa o número final de grupos. A definição do valor de  $k$  pode ser feita usando o método do cotovelo. O método do cotovelo é uma heurística usada para determinar o número de *clusters* de um conjunto de dados. O método consiste em traçar a pontuação de distorção em função do número de *clusters* e escolher o cotovelo da curva como o número de *clusters* a utilizar. A pontuação de distorção é uma métrica que calcula a soma das distâncias quadradas de cada ponto até ao centro atribuído. A intuição por trás da curva do cotovelo é que a variação da pontuação de distorção muda rapidamente até o número de grupos presentes nos dados e então diminui levando a uma formação de cotovelo no gráfico. O ponto de cotovelo é o número de *clusters*, ou seja, o melhor  $k$  a utilizar. Existem vários métodos possíveis de inicialização do *Kmeans* como, por exemplo, o método *Forgy* e o *Random Partition*. Para encontrar os melhores protótipos para os grupos, é necessário uma medida de distância. Medidas populares são a distância euclidiana, de *Manhattan* e *Minkowsky* [20].

Este algoritmo não apresenta um bom desempenho em dados com ruído, com grupos de tamanhos diferentes ou formas não globulares.

### 2.3.3 Getis-Ord $G_i^*$ do ArcGIS

A estatística *Getis-Ord  $G_i^*$*  é utilizada na ferramenta de análise de *hotspots* do software ArcGIS e tem sido utilizada para a análise espacial em alguns trabalhos [35, 40]. Esta ferramenta permite identificar grupos espaciais estatisticamente significativos de valores altos, *hotspots*, e de valores baixos, *coldspots*.

Esta estatística é calculada para cada recurso do conjunto de dados, produzindo uma pontuação  $Z$  e um valor  $p$  para cada um. Estes resultados indicam onde os recursos se agrupam geograficamente. Esta estatística funciona examinando cada recurso no contexto de eventos vizinhos. Para ser um *hotspot* estatisticamente significativo, um recurso tem de ter um valor alto e estar cercado por outros recursos com valores altos. A soma local de um recurso e dos seus vizinhos é comparada proporcionalmente à soma de todos os recursos. Quando a soma local é muito diferente da soma local esperada e quando essa diferença é muito grande para resultar de hipótese aleatória, obtém-se uma pontuação  $Z$  estatisticamente significativa [2].

A hipótese nula utilizada por esta ferramenta é a *Randomity Spatial Complete*. As pontuações  $Z$  e os valores  $p$  retornados informam se se pode rejeitar essa hipótese nula ou não [3]. Para rejeitar a hipótese nula, deve-se seleccionar um grau de confiança.

O valor  $p$  é a probabilidade de que o padrão espacial observado tenha sido criado por um processo aleatório. Quando o valor  $p$  é muito pequeno, significa que é muito improvável que o padrão espacial observado seja o resultado de processos aleatórios, podendo-se rejeitar a hipótese nula. A pontuação  $Z$  é um desvio padrão. Ambos os resultados estão associados à distribuição normal. Uma pontuação  $Z$  alta e um valor  $p$  pequeno para um recurso indica um agrupamento espacial de valores altos. Uma pontuação  $Z$  negativa

baixa e um valor  $p$  pequeno indica um agrupamento espacial de valores baixos. Uma pontuação  $Z$  próxima de zero indica que não existe agrupamento espacial [3].

#### 2.3.4 Space-time Scan Statistics

Este é um método de *clustering* espaço-temporal e tem sido usado na literatura [28, 29] através do *software SatScan*. O *SaTScan* faz uso das estatísticas de varredura espacial, temporal e espaço-temporal para análise de dados espaço-temporais. Os trabalhos relacionados [28, 29] utilizaram a estatística de varredura usando o modelo de *Poisson*.

As estatísticas de varredura espacial são projectadas para detectar grupos significativos de eventos, movendo uma janela circular de raio variável num conjunto de dados e registando os círculos que exibem números estatisticamente significativos de eventos. Um grande número de janelas circulares são geradas e cada uma cobre um conjunto diferente de eventos. Portanto, cada janela circular é um possível candidato a grupo [24].

As estatísticas de varredura espaço-temporal empregam um método semelhante. Estas estatísticas analisam os eventos usando informações espaciais e temporais. São utilizados cilindros em três dimensões, ou seja, duas dimensões espaciais e uma dimensão temporal. A base do cilindro representa uma janela circular bidimensional no espaço geográfico e a altura do cilindro representa o intervalo de tempo. Semelhante às estatísticas de varredura espacial, o raio das bases e as alturas do cilindro podem variar, de modo que as janelas cilíndricas cubram cada combinação possível de localização e tamanho no espaço tridimensional espaço-tempo [24, 28].

O tamanho das janelas espaciais e temporais varia de zero até ao máximo, ou seja, até 50% da população total em risco e 50% do período temporal. A presença de grupos espaciais e temporais de risco elevado é avaliada usando um teste de razão de verossimilhança. Este teste determina se o número de casos observados excede o número esperado de casos em cada janela de varredura com base no número total de casos e no tamanho da população observado em toda a área geográfica durante um período de tempo. A proporção do observado para o esperado denota o risco dentro da janela e o risco relativo compara o risco dentro da janela com o risco fora da janela. Usando o modelo de *Poisson*, o teste da razão de verossimilhança é realizado sob a hipótese nula de aleatoriedade espacial e temporal completa, ou seja, o risco dentro da janela é igual ao risco fora da janela. A janela com a taxa máxima de verossimilhança é denotada como o grupo mais provável e os grupos secundários são identificados e classificados de acordo com suas taxas de verossimilhança. A significância estatística dos grupos identificados é determinada usando uma abordagem de *Monte Carlo* [24, 28].

#### 2.3.5 DBSCAN

O algoritmo **DBSCAN**, apesar de não ter sido utilizado nos trabalhos relacionados, é muito conhecido na comunidade, pois consegue lidar com ruído, não necessita da especificação do número de grupos e consegue identificar grupos com diferentes formas, ao



contrário do *K-means*. No entanto, o **DBSCAN** falha ao encontrar grupos com densidades diferentes [39].

Definindo a vizinhança  $N_\epsilon$  de cada ponto como o conjunto de pontos à distância  $\epsilon$ , um ponto  $p$  é um ponto central se o número de pontos em  $N_\epsilon$  de  $p$  é pelo menos igual ao parâmetro *minPts*. Um ponto  $q$  é alcançável a partir de  $p$  se  $p$  é um ponto central e  $q$  está na vizinhança  $N_\epsilon$  de  $p$  ou na vizinhança  $N_\epsilon$  de qualquer ponto central que está acessível a partir de  $p$ . Esta definição recursiva significa que  $q$  é acessível a partir de  $p$  se houver um caminho de pontos centrais alcançáveis de  $p$  a  $q$ . O algoritmo funciona da seguinte forma: para cada ponto  $p$ , se o número de pontos na vizinhança  $N_\epsilon$  de  $p$  for menor que *minPts*, presume-se que  $p$  seja ruído. Caso contrário, um grupo é criado para  $p$ , que é um ponto central, e todos os vizinhos de  $p$  são adicionados ao grupo. Se qualquer vizinho de  $p$  for um ponto central pertencente a outro grupo, os grupos serão fundidos [16].

Tal como aconteceu com o **KDE**, foi proposta uma extensão do **DBSCAN** para lidar com dados espaço-temporais. Essa proposta foi o *Spatial-temporal Density-based spatial clustering of applications with noise* (**ST-DBSCAN**). Este algoritmo requer quatro parâmetros,  $\epsilon_s$ ,  $\epsilon_t$ , *MinPts* e  $\Delta_\epsilon$ . O  $\epsilon_s$  é o parâmetro de distância para atributos espaciais e o  $\epsilon_t$  é o parâmetro de distância para atributos não espaciais. Qualquer métrica de distância usada no **DBSCAN** pode ser usada para estes dois parâmetros. O *MinPts* é, tal como no **DBSCAN**, o número mínimo de pontos para formar um grupo. Se uma região é densa, então deve conter mais pontos que o valor de *MinPts*. O parâmetro  $\Delta_\epsilon$  é usado para impedir a descoberta de grupos combinados devido às pequenas diferenças nos valores não espaciais dos vizinhos locais [5].

O algoritmo começa com o primeiro ponto  $p$  e recupera todos os pontos acessíveis pela densidade de  $p$  em relação a  $\epsilon_s$  e  $\epsilon_t$ . Se  $p$  é um ponto central, um grupo é formado. Se  $p$  for um ponto não central, nenhum ponto poderá ser atingido pela densidade de  $p$  e o algoritmo visitará o próximo ponto. O processo é repetido até que todos os pontos tenham sido processados [5].

### 2.3.6 SNN

O algoritmo **SNN** é capaz de identificar grupos de diferentes formas, tamanhos e densidades. Também tem a capacidade de lidar com ruído, tal como o **DBSCAN**. Este algoritmo faz uso, na sua última fase, do algoritmo **DBSCAN**. A vantagem deste algoritmo fase ao **DBSCAN** é conseguir identificar grupos com diferentes densidades, ao contrário do **DBSCAN** que apenas considera grupos de densidade semelhante. A figura 2.1 ilustra as diferenças dos resultados do **DBSCAN** e **SNN** quando executados com um conjunto de dados que contém grupos de diferentes densidades.

Em conjuntos de dados de grande dimensão, a noção euclidiana tradicional de densidade, que é o número de pontos por unidade de volume, não faz sentido. Então o algoritmo **SNN** usa diferentes medidas de similaridade em termos de KNNs (*k-nearest neighbors*) e depois define a densidade em termos dessa semelhança. São necessário três parâmetros

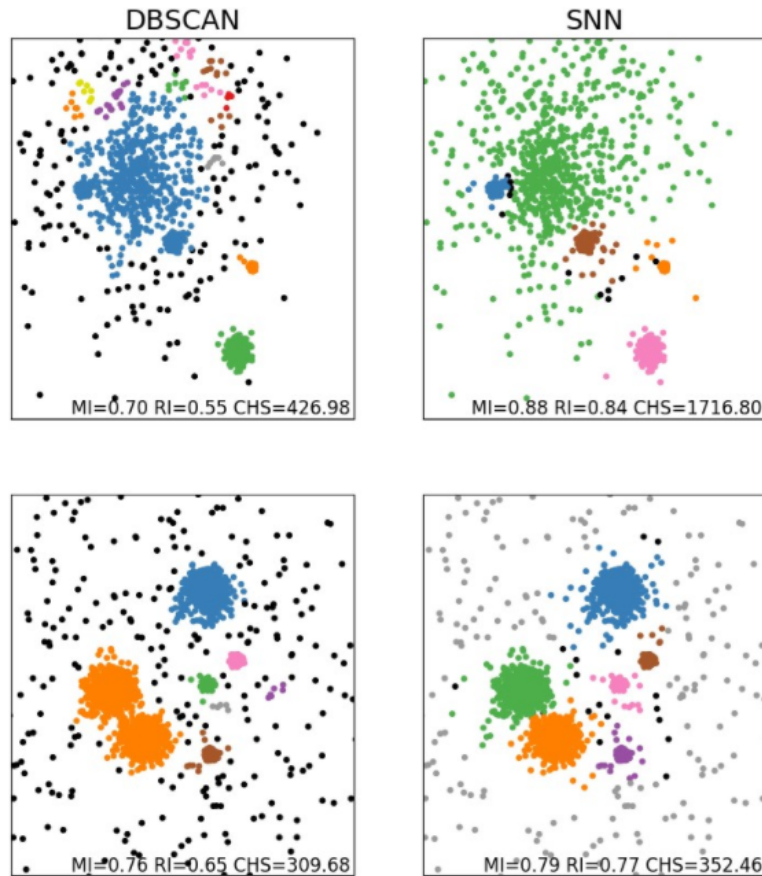


Figura 2.1: Diferença de resultados obtidos entre o algoritmo de *clustering* DBSCAN e o SNN para o mesmo conjunto de dados [15].

para o SNN: o  $k$  (número de vizinhos mais próximos que devem ser identificados para cada ponto),  $\epsilon$  (número mínimo de vizinhos que dois pontos devem partilhar para serem considerados semelhantes entre si) e  $MinPts$  (densidade mínima que um ponto deve ter para ser considerado um ponto central) [34]. As etapas do SNN são as seguintes [14]:

1. Calcular uma matriz de similaridade, que corresponde a um grafo de similaridade onde os pontos são nós e as arestas têm pesos relativos às semelhanças entre os pontos;
2. Manter apenas as  $k$  ligações mais fortes do grafo de similaridade;
3. Construir o grafo do vizinho mais próximo partilhado a partir da matriz de similaridade com as ligações mais fortes;
4. Executar o algoritmo de DBSCAN.
  - a) Encontrar as densidades de todos os pontos, usando o parâmetro  $\epsilon$ , ou seja, contar quantas ligações do grafo o ponto tem com valor igual ou superior a  $\epsilon$ . Esta é a densidade SNN dos pontos;



- b) Encontrar os pontos centrais, usando o parâmetro *MinPts*, isto é, todos os pontos que possuem uma densidade SNN maior que *MinPts*;
- c) Formar os grupos a partir dos pontos centrais, se dois pontos centrais estiverem dentro de um raio,  $\epsilon$ , um do outro são colocados no mesmo grupo;
- d) Descartar todos os pontos de ruído;
- e) Associar a todos os pontos sem ruído não centrais um grupo.

Existe uma proposta de extensão do algoritmo SNN denominada *Spatial-temporal Shared Nearest Neighbor (ST-SNN)*. Este foi proposto para agrupar dados espaço-temporais. O ST-SNN precisa de quatro parâmetros de entrada: o  $k$  (número de eventos numa vizinhança espaço-temporal), o  $k_t$  (limite usado para determinar se dois eventos são directamente alcançáveis espaço-temporalmente), o *MinPts* (para identificar um núcleo espaço-temporal) e o  $\Delta T$  (janela de tempo para definir uma vizinhança espaço-temporal). Os passos do algoritmo são os seguintes [26]:

1. Identificar os  $k$  vizinhos espaço-temporais para cada evento;
2. Pesquisar os vizinhos espaço-temporais partilhados mais próximos para cada evento;
3. Calcular a densidade espaço-temporal de cada evento e detectar núcleos espaço-temporais;
4. Expandir grupos espaço-temporais. Seleccionar um núcleo espaço-temporal e usar uma estratégia recursiva para adicionar todos os eventos que são alcançáveis e conectivos para construir um grupo.
5. Identificar eventos de ruído.

### 2.3.7 Conclusões

O *clustering* é um processo de particionar um conjunto de dados num conjunto de subclasses significativas, chamados grupos. O conjunto de dados é armazenado em grupos para extrair padrões dele [21].

Relativamente a técnicas de *clustering* para detectar padrões, grande parte da literatura faz uso do KDE para *clustering* espacial. Existe uma variante do KDE para *clustering* espaço-temporal denominado STKDE que também é utilizado. Outro algoritmo aplicado é o *K-means*, este algoritmo serviu tanto para agrupar dados geograficamente como para séries temporais. A estatística *Getis-Ord Gi\** da ferramenta *Hot Spot Analysis* do software ArcGIS é usada para a detecção de *hotspots*. Por fim, a estatística *Space-Time Scan Statistic* também foi usada em alguns trabalhos relacionados. Foram analisados ainda outros algoritmos como o DBSCAN e o SNN. O DBSCAN é um bom algoritmo de agrupamento para quando se quer considerar ruído nos dados, no entanto, o SNN apresenta uma vantagem sobre o DBSCAN, pois consegue agrupar grupos de diferentes densidades, o que não acontece no DBSCAN.

Existem muitos esforços da comunidade de aprendizagem automática para o desenvolvimento de algoritmos que tenham em conta a componente espacial e temporal dos dados [39]. Já existem algoritmos que têm as duas componentes em conta, por exemplo, o *STKDE*, *ST-DBSCAN* e *ST-SNN* [5, 6, 26].

## 2.4 Técnicas de Visualização e Interação de Dados

Nesta secção serão apresentadas as técnicas de visualização que foram utilizadas nos trabalhos relacionados. Também serão apresentadas outras técnicas que, apesar de não terem sido utilizadas nos trabalhos relacionados, existem, são relevantes e podem ser aplicadas em dados espaço-temporais. Por fim, serão apresentadas formas de interação com visualizações espaço-temporais.

### 2.4.1 Temporal

A análise temporal geralmente é mostrada num simples gráfico de linhas. No entanto esta visualização quando utilizada no contexto de análise temporal ao nível horário não é capaz de ilustrar a continuidade cronológica do tempo, isto é, não é capaz de captar a ciclicidade existente [4]. Outro método de visualização que pode ser aplicado a dados temporais são os gráficos de radar, um exemplo está presente na Figura 2.2. Estes gráficos são bidimensionais e são projectados para mostrar uma ou mais séries de valores sobre várias variáveis quantitativas. Cada variável tem o seu próprio eixo e todos os eixos são unidos no centro da figura. Ao contrário dos gráficos de linhas estes gráficos são capazes de ilustrar a continuidade e a ordem cronológica. Com o gráfico de radar é fácil detectar semelhanças através das formas. Muitos dos trabalhos relacionados encontrados usam o gráfico de radar para fazer a análise temporal em diferentes granularidades [4, 11, 40].

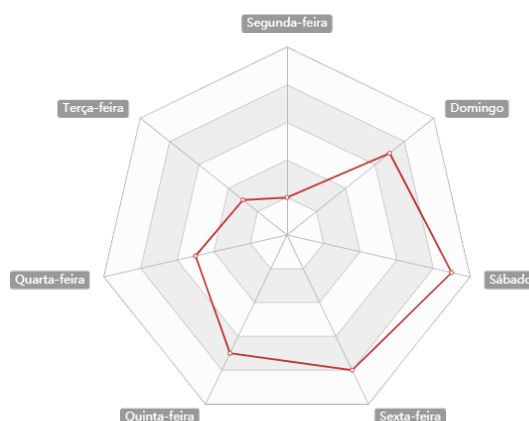


Figura 2.2: Exemplo de gráfico de radar.

Outra visualização que está presente na literatura é o mapa de calor, um exemplo está presente na Figura 2.3. Esta visualização foi utilizada numa perspectiva de análise de padrões temporais [35]. O mapa de calor é uma representação gráfica de dados em que os valores individuais contidos numa matriz são representados como cores. Por norma, cores mais escuras denotam valores maiores e cores mais claras valores menores.

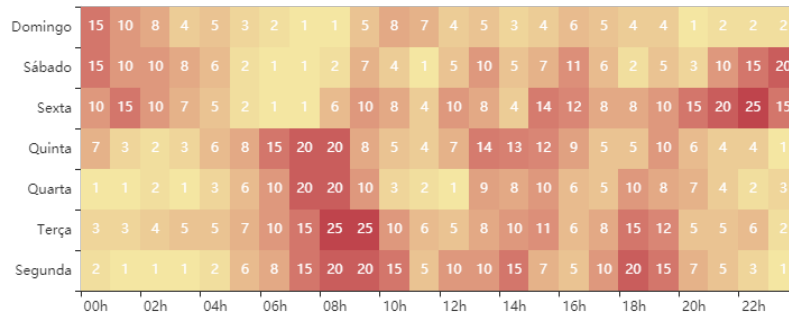


Figura 2.3: Exemplo de mapa de calor.

### 2.4.2 Espacial

A análise espacial é frequentemente feita através do mapeamento das coordenadas geográficas num mapa. Acontece em muitos casos que o volume de eventos mapeados cria problemas de visualização ao nível do ponto, onde várias localizações de eventos aparecem como uma única ocorrência, cada ponto é posicionado um em cima do outro acabando sobrepostos [12]. A Figura 2.4a representa um mapa de pontos, onde cada ponto é um evento. A Figura 2.4b representa um mapa de densidade. Os mapas de densidade podem gerados por técnicas semelhantes ao KDE. Com os exemplos presentes na Figura 2.4 é possível verificar as restrições da visualização de eventos em mapas de pontos onde existe um grande volume de dados comparativamente com o mapa de densidade que apresenta uma visualização melhor para compreender a distribuição espacial dos eventos.

A utilização de mapas de densidade foi largamente utilizada nos trabalhos relacionados para a análise espacial [4, 8, 11, 12].

Outra técnica de visualização também usada na literatura [4, 11, 12, 40] é o mapa coroplético, presente na Figura 2.5. Um mapa coroplético é uma mapa temático e representa um mapa espacial filtrado usando filtros espaciais, de tamanho variável, sem sobreposição com formas seleccionadas de regiões políticas ou administrativas disponíveis. Exibe áreas geográficas divididas que são coloridas em relação a uma variável numérica, permitindo estudar como essa variável evolui ao longo do território [12]. No entanto esta visualização apresenta algumas desvantagens. O uso de limites administrativos pode colocar restrições artificiais nos dados e isto influencia a interpretação dos resultados. É necessário haver normalização da variável, pois não se pode comparar números brutos entre regiões de tamanhos diferentes.

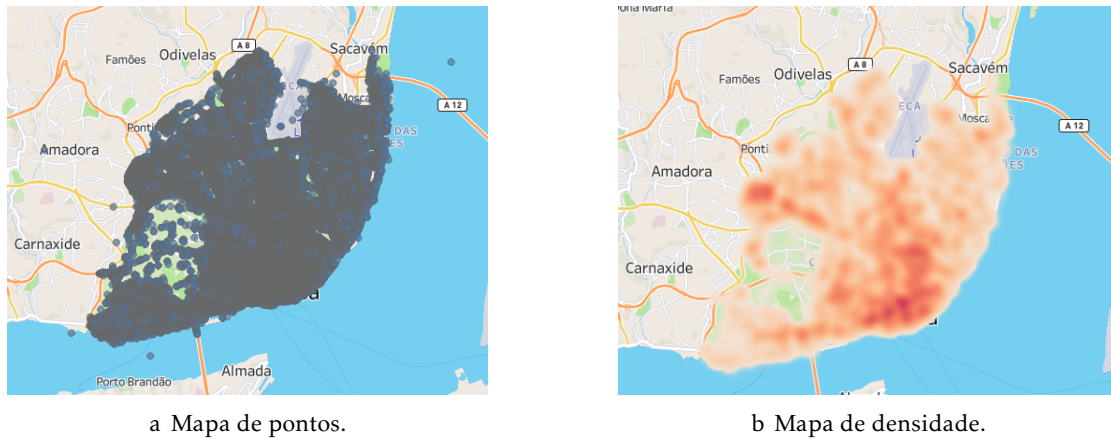


Figura 2.4: Mapa de pontos e mapa de densidade.

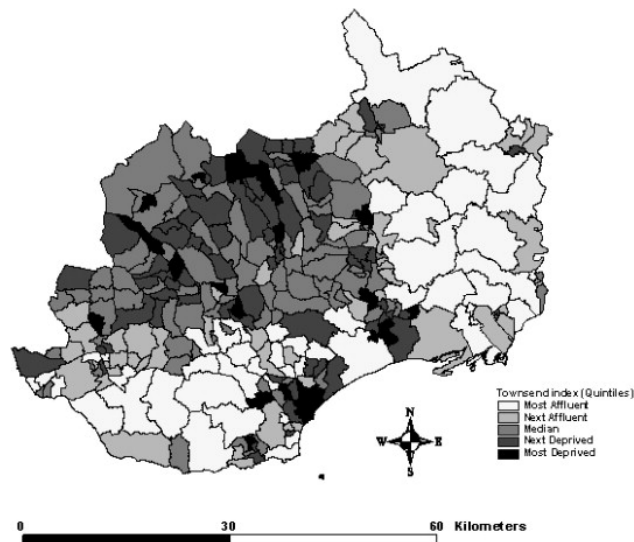


Figura 2.5: Exemplo de mapa coroplético [12].

Existe outra técnica de visualização que permite ultrapassar o problema do uso de limites administrativos. Esta técnica é o mapa de *hexabins*. Nesta visualização todas as regiões do mapa são representadas como hexágonos. A área é dividida em uma infinidade de hexágonos e o número de pontos em cada hexágono é contado e representado usando um gradiente de cores. Aqui nenhum limite no mapa é necessário.

### 2.4.3 Espaço-Temporal

Para a análise das componentes espaciais e temporais, simultaneamente, um método muito utilizado nos trabalhos relacionados [4, 8, 11] foi o *comap*. A vantagem do *comap* é a sua capacidade de ilustrar todo o período em estudo numa única visualização [11]. Onde existem inúmeras mudanças nos padrões, uma abordagem de visualização sinóptica permite que seja feita uma análise comparativa superior entre as dimensões temporal e espacial de toda a cobertura temporal, em contraste com a clássica animação do mapa,

onde cada período de tempo é representado como uma única visualização [1]. Esta visualização baseia-se numa técnica conhecida como *co-plot* que é usada para examinar a relação entre um par de variáveis que são condicionados usando uma terceira variável [11]. A Figura 2.6A mostra seis painéis condicionados usando a variável da hora. Cada painel corresponde a um subconjunto, intervalo de tempo. Existem duas regras para os subconjuntos, o intervalo de cada subconjunto deve ter alguma sobreposição com cada subconjunto adjacente e cada subconjunto deve conter aproximadamente o mesmo número de observações. A Figura 2.6B mostra a interpretação de um *comap* onde os dados são condicionados a duas variáveis. No último *comap* os intervalos para a segunda variável não se sobrepõem, porque o período do dia é uma variável categórica e não contínua como hora e, portanto, deve ser tratado de maneira diferente [6, 11].

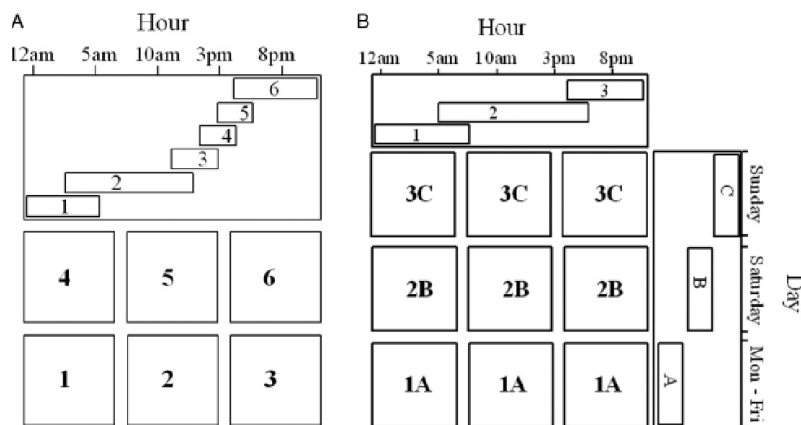
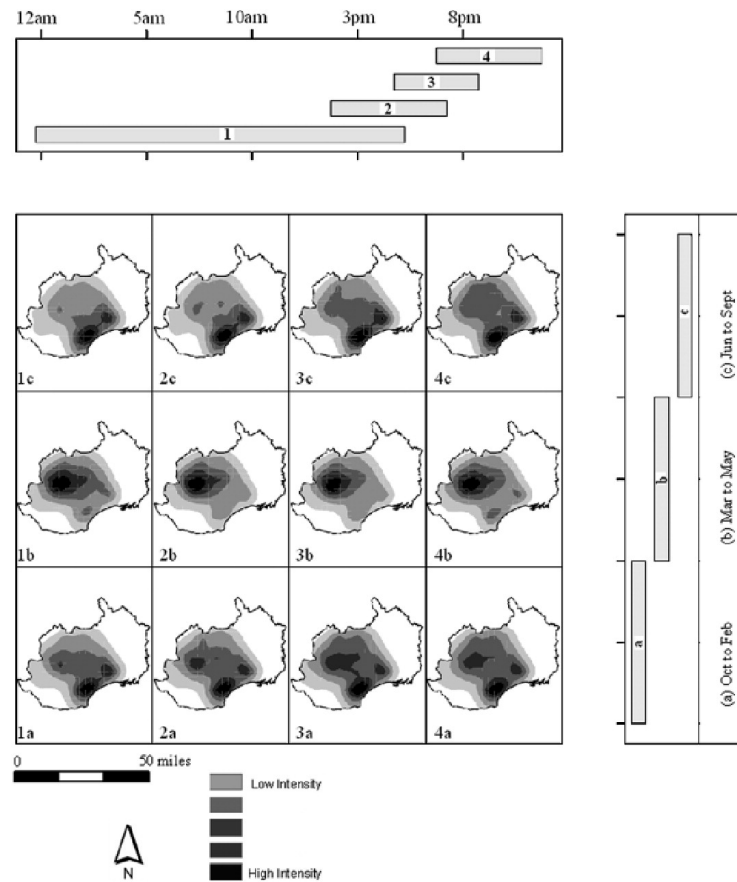


Figura 2.6: Estrutura do *comap* [11].

A Figura 2.7 mostra a utilização dos *comaps* em conjunto com o resultado do KDE. Esta abordagem foi utilizada em grande parte dos trabalhos relacionados [4, 8, 11].

Actualmente têm surgido técnicas de visualização tridimensionais. Uma dessas técnicas é a isosuperfície [6], presente na Figura 2.8. Esta abordagem permite visualizar padrões de eventos durante períodos de tempo numa única visualização. A abordagem é considerar cada evento como um ponto no espaço-tempo  $(x, y, t)$ . A isosuperfície fornece uma maneira de visualizar funções de densidade, como o KDE, no espaço tridimensional. Para funções de três variáveis, uma isosuperfície é o equivalente a uma curva de nível para funções de duas variáveis. Ou seja, é uma superfície bidimensional embutida no espaço tridimensional que une pontos com o mesmo valor (valor  $v$ ) de uma função aplicada aos três argumentos representados pelos pontos. Normalmente, para funções de duas variáveis, as curvas de nível mostram várias linhas para vários valores de  $v$ . No entanto, para funções de três variáveis existem dificuldades ao tentar mostrar mais de um valor de  $v$ . Superfícies de um valor de  $v$  envolvem outras superfícies no espaço tridimensional, tornando-as invisíveis. Uma abordagem possível para mostrar mais de uma isosuperfície na mesma imagem é desenhar uma superfície externa com um grau de transparência,

Figura 2.7: Exemplo do *comap* [11].

de modo que uma segunda superfície envolvida nela seja visível. Infelizmente, os resultados desta visualização não são bons, em muitos casos as isosuperfícies são difíceis de identificar [6].

Em trabalhos relacionados mais recentes outra visualização tridimensional usada foi o cubo espaço-tempo [28, 29]. Um cubo espaço-tempo é tridimensional e consiste em duas dimensões horizontais do espaço e uma dimensão vertical do tempo. Os pontos no espaço tridimensional representam as posições de um objecto em diferentes momentos do tempo. Podem ser usadas linhas que ligam os pontos correspondentes a momentos consecutivos. A exibição do cubo espaço-tempo, para além de permitir visualizar trajetórias de movimento, permite a exploração da velocidade do movimento. Nesta representação, segmentos levemente inclinados indicam movimento rápido, isto é, longa distância percorrida em pouco tempo, enquanto segmentos íngremes correspondem a movimentos lentos. Linhas verticais ocorrem quando um objecto permanece por algum período de tempo no mesmo local. Portanto, o cubo expõe certos padrões que fornecem informações valiosas sobre características do movimento do objecto [17].

Esta visualização pode ser usada para demonstrar resultados de técnicas de *clustering* espaço-temporais. Assim, esta visualização é útil na diferenciação de grupos estáveis, bem como na identificação de difusão geográfica e movimentação de grupos [29].



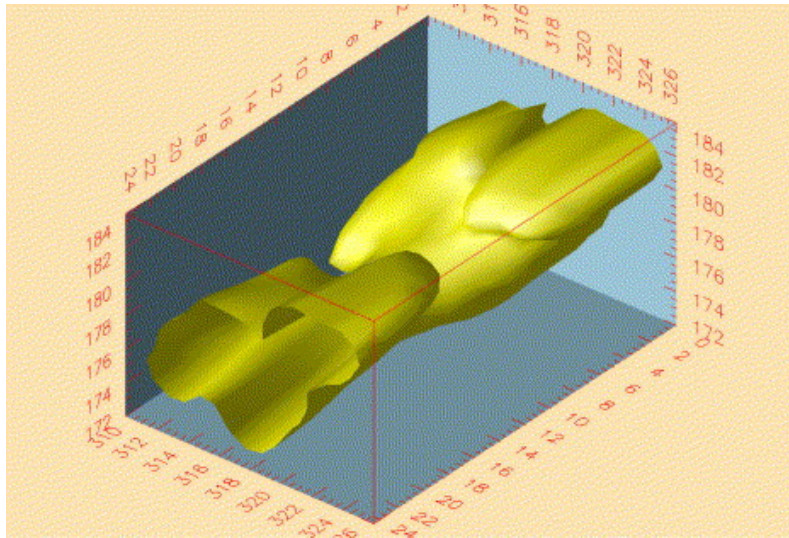


Figura 2.8: Exemplo de uma isosuperfície [6].

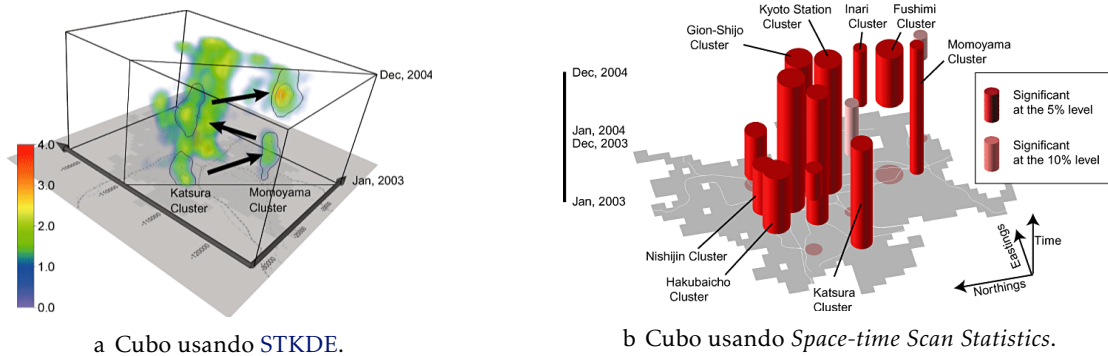


Figura 2.9: Exemplo da utilização do cubo espaço-tempo para visualizar grupos [29].

A Figura 2.9 demonstra a utilização desta visualização para a análise espácio-temporal. Na Figura 2.9a o cubo espaço-tempo foi utilizado para visualizar o resultado do algoritmo espácio-temporal STKDE. Na Figura 2.9b o cubo espaço-tempo foi utilizado para visualizar o resultado do *Space-time Scan Statistics*.

#### 2.4.4 Interações com visualizações espácio-temporais

Em conjunto com a representação visual, as interações permitem ao utilizador formar modelos mentais de forma implícita das relações existentes dentro dos dados, da identificação de padrões e da busca de hipóteses de análise. Existem várias finalidades da interatividade como: tornar os dados mais envolventes, controlar o que pode ser visível na interface, aumentar a quantidade e variedade de perspectivas analíticas para satisfazer diferentes necessidades, apresentar os dados em porções flexíveis. Essa flexibilidade é particularmente importante ao explorar conjuntos de dados volumosos, onde diminuir o seu nível de complexidade é vital para seu entendimento e análise [33].

Conjuntos de dados espácio-temporais são, frequentemente, volumosos e desafiantes

tanto para analisar como para exibir. A articulação das dimensões temporais e espaciais é muito poderosa, pois permite aos utilizadores interpretar a evolução dos eventos no tempo e no espaço. O trabalho presente em [33] explora quais são as especificidades do uso de interações em visualizações espaço-temporais aplicando uma taxonomia de interação genérica. As interações dessa taxonomia e respectivas descrições estão presentes na tabela 2.1. Essa taxonomia foi aplicada a vários exemplos de visualizações espaço-temporais, casos de estudo, e foi realizada uma análise sistemática das técnicas de interação aplicadas nessas visualizações. Foram detetadas situações onde as interações eram inviáveis no contexto espaço-temporal. Portanto, foi desenvolvida uma nova taxonomia a partir da taxonomia genérica. Esta nova taxonomia assume, como foco conceptual, a seguinte estrutura:

- Tema - refere-se às características, à temática dos dados;
- Tempo - refere-se às características temporais dos dados;
- Espaço - refere-se às características geográficas dos dados.

Tabela 2.1: Taxonomia de interação de visualização proposta por Figueiras [33].

Tipo de interação	Descrição da interação
Filtrar	Exibir apenas dados nos quais estou interessado.
Selecionar	Marcar ou rastrear itens nos quais estou interessado.
Abstrato/Elaborar	Ajustar o nível de abstração dos dados.
Visão geral e explorar	Visão geral primeiro, zoom e filtro, depois detalhes sob demanda.
Conetar/Relacionar	Mostrar como estes dados estão relacionados.
Reconfigurar	Dar uma disposição diferente dos dados.
Codificar	Dar uma representação diferente dos dados.
História	Permitir refazer as etapas que eu realizei durante a exploração.
Extração de características	Permitir extrair os dados em que estou interessado.
Participação/Colaboração	Permitir contribuir com dados.
Gamificação	Mostrar os dados de uma forma mais lúdica.

Foi feita a tentativa de decomposição de cada tipo de interação nas três dimensões da estrutura, Tema, Tempo e Espaço. No entanto, nem todos os tipos de interações podem ser decompostas nas três dimensões [33]. O resultado dessa decomposição, ou seja, a nova taxonomia criada está presente na tabela 2.2. As interações que podem ser decompostas nas três dimensões são a filtragem, seleção, conectar/relacionar, reconfigurar e codificar. A interação de ampliar apenas pode ser decomposta no tempo e no espaço. As interações de detalhes sob demanda, ligar, extração de características, participação/colaboração e gamificação apenas podem ser decompostas pela componente temática dos dados. Por fim, as interações visão global e história não são adequadas para decomposição em nenhuma das três dimensões.

Relativamente às filtragens, estas podem ser executadas nas três dimensões, Tema, Tempo e Espaço, sendo que para o Tema, isto significa a possibilidade de visualizar apenas



objetos com valores ou características específicas. No Tempo, permite a escolha de um momento ou intervalo de tempo. A dimensão Espaço pode ser filtrada para visualizar resultados de uma área geográfica de interesse [33].

A técnica de interação seleção também pode ser executada nas três dimensões. No Tema, a seleção engloba a possibilidade de marcar e rastrear itens para fazer a distinção dos restantes dados. No Tempo, é a escolha de um momento preciso, por exemplo, para contrastar um ponto no tempo. Por fim, a seleção no Espaço é quando uma área específica é destacada do resto [33].

A interação de conectar/relacionar no contexto espaço-temporal, no Tema é quando o utilizador pode comparar e pesquisar relações entre dois eventos diferentes. Quando se compara valores para um mesmo objeto em vários momentos, estabelece-se um conectar/relacionar no Tempo e a comparação de dados de dois locais distintos é conectar/relacionar na dimensão Espaço [33].

A interação de reconfiguração no Tema fornece uma nova organização das informações sobre o tópico, permitindo novas perspectivas de análise. No Tempo, uma reconfiguração poderia ser a mudança de uma ordem cíclica para uma escala linear, exibindo pontos de tempo ou intervalos de tempo. No Espaço, uma reconfiguração seria a reorganização da posição geográfica de um local [33].

A técnica de interação codificação no Tema pode permitir a variação dos componentes temáticos, por exemplo, alterar as cores dos elementos gráficos exibidos para melhorar a legibilidade da exploração de dados. No Tempo, a codificação, por exemplo, permite ao utilizador alternar entre calendários. No Espaço, oferece uma aparência visual diferente dos componentes geográficos, por exemplo, a alternância entre diferentes projeções de mapas [33].

O ampliar só é possível decompor em duas das três dimensões. Na dimensão Tempo, o ampliar dá a sensação de aproximação na escala temporal ao esticar ou retrair a quantidade de detalhes. O ampliar no Espaço influencia o nível de detalhe da informação espacial exibida na visualização do mapa, aumentar e diminuir o *zoom* [33].

O detalhes sob demanda são materializados por meio da exibição de outros detalhes informativos, por exemplo, elementos textuais/narrativos. A interação de ligar fornece caminhos internos para objetos externos para conteúdo relacionado, como vídeos. A extração de características, como conjuntos de dados ou imagens também estão sempre sujeitos aos dados temáticos. Todas estas interações só são decompostas no Tema [33].

Técnicas de interação como a participação/colaboração e gamificação, que são mais complexas, estão restritas à dimensão Tema. A participação/colaboração toma forma quando a visualização inclui recursos fornecidos pelos colaboradores. Essas contribuições são assim mescladas na visualização, contribuindo para o Tema. A gamificação pode ser materializada na forma de Quizzes, pequenos desafios de adivinhar quem, o quê, onde e quando [33].

Algumas técnicas de interação são intuitivas e frequentemente aplicadas no contexto

de visualização espaço-temporal, como a filtragem no tempo, o ampliar no espaço e exibição de detalhes sob demanda. As interações mais complexas, como a participação/colaboração e gamificação raramente são incorporadas, pois são trabalhosas de implementar [33]. A generalização do uso da técnica conectar/relacionar seria capaz de reforçar ações de comparação, análises de tendências e padrões.

### 2.4.5 Conclusões

Os cenários exploratórios consideram a visualização como uma ferramenta de descoberta, onde é necessário fornecer meios para que os usuários encontrem as respostas às suas questões [33].

O primeiro passo para entender e explorar um conjunto de dados para obter informações relevantes é uma boa escolha de técnicas de visualização de dados, gráficos e diagramas [23]. Relativamente às técnicas de visualização, existem visualizações bidimensionais que conseguem ilustrar um determinado período de tempo numa única visualização, os *comaps* [4, 8, 11] e, mais recentemente, visualizações tridimensionais também têm sido adoptadas para detectar padrões espaço-temporais, o cubo espaço-tempo [28, 29]. Estas últimas visualizações são usadas para apresentar o resultado dos algoritmos de *clustering* espaço-temporais. É possível fazer a análise da componente temporal e espacial em separado com recurso a, como por exemplo, gráficos de radar [4, 11], mapas de calor [35] e mapas de densidade [12]. No entanto, esta abordagem não consegue captar o mesmo que uma análise das componentes em simultâneo, por exemplo, não é capaz de captar a evolução geográfica dos eventos ao longo do tempo.

Relativamente às interações, foi feito um levantamento de quais os tipos de interações existentes para dados genéricos e também para dados espaço-temporais. Concluiu-se que algumas das interações funcionam tanto no contexto genérico como em todo o contexto espaço-temporal (nas três dimensões dos dados, o Tema, o Tempo e o Espaço), como é o exemplo da filtragem, seleção, conectar/relacionar, reconfigurar e codificar. No entanto, existem alguns tipos de interações que não podem ser aplicadas a todas as dimensões dos dados espaço-temporais, mas sim apenas à sua componente temática, como a extração de características, participação/colaboração e gamificação. Por fim, interações do tipo visão global e história não são adequadas para decomposição em nenhuma das dimensões presentes nos dados espaço-temporais.

Para finalizar, concluiu-se que existe uma lacuna na literatura relativamente ao processo de *design* e documentação de *dashboards*. O IFML - *The Interaction Flow Modeling Language*<sup>1</sup> serve para expressar o conteúdo, a interação do utilizador e o comportamento *front-end* em sistemas de *software*. Esta linguagem é essencialmente utilizada para o desenvolvimento de páginas *web*. Portanto, durante a pesquisa não foi encontrada nenhuma linguagem gráfica que apoie o processo de *design* e de documentação de *dashboards*, um conjunto de visualizações e respetivas interações entre elas.

---

<sup>1</sup><https://www.ifml.org/>

Tabela 2.2: Taxonomia de interação de visualizações espaço-temporais [33].

Aplicação	Tipo de interação	Dimensão	Descrição da interação
Espácio-temporal	Filtrar	Tema	Exibir apenas itens com as propriedades em que estou interessado.
		Tempo	Exibir apenas pontos de tempo ou intervalos nos quais estou interessado.
		Espaço	Exibir apenas locais nos quais estou interessado.
	Selecionar	Tema	Marcar ou rastrear itens com as propriedades nas quais estou interessado.
		Tempo	Marcar ou acompanhar os pontos ou intervalos de tempo nos quais estou interessado.
		Espaço	Marcar ou rastrear locais nos quais estou interessado.
	Ampliar	Tempo	Variar o nível de abstração de um intervalo de tempo.
		Espaço	Ampliar ou reduzir uma área.
	Conectar / Relacionar	Tema	Relacionar dois ou mais itens de temas diferentes.
		Tempo	Relacionar dois ou mais pontos ou intervalos de tempo.
	Reconfigurar	Espaço	Relacionar duas ou mais localizações.
		Tema	Dar uma disposição diferente dos dados.
		Tempo	Dar um arranjo diferente da estrutura temporal.
	Codificar	Espaço	Dar uma disposição diferente dos locais.
		Tema	Dar um código visual diferente para os componentes temáticos.
		Tempo	Dar um código visual diferente para a estrutura temporal.
Genérico	Visão Global	Espaço	Dar um código visual diferente para os componentes geográficos.
		Primeiro	Dar uma visão geral de todas as ocorrências primeiro.
		Último	Dar uma visão geral de todas as últimas ocorrências.
	Detalhes Sob demanda	Sob demanda	Dar uma visão geral de todas as ocorrências a qualquer momento.
		Tema	Mostrar os detalhes de um item específico sob demanda.
	Ligar	Tema	Levar a informações complementares.
	História	-	Permitir refazer os passos que foram feitos na exploração.
	Extração de características	Tema	Permitir extrair dados nos quais estou interessado.
	Participação / Colaboração	Tema	Permitir contribuir com dados.
	Gamificação	Tema	Mostrar os dados de uma forma mais lúdica.



## TRABALHO REALIZADO

Neste capítulo será apresentado todo o trabalho que foi desenvolvido durante o desenvolvimento desta dissertação.

Este capítulo contém, na seção 3.1, a abordagem que foi seguida para atingir os objetivos desta dissertação. Serão apresentadas quais as tecnologias/ferramentas utilizadas para desenvolver o trabalho na seção 3.2, nomeadamente, para a criação de visualizações e *dashboards*, para processamento de dados, para o desenvolvimento de programas e para a criação de diagramas. Também serão apresentados os conjuntos de dados utilizados e os respetivos processamentos de que foram alvo na seção 3.3. Serão discutidos os agrupamentos realizados durante a dissertação, nomeadamente, agrupamentos temporais e espaciais na seção 3.4. Este capítulo também contém uma apresentação detalhada dos *dashboards* que foram desenvolvidos na seção 3.5. Por fim, também serão apresentadas algumas análises e descobertas feitas utilizando as ferramentas desenvolvidas na seção 3.6.

### 3.1 Abordagem

Primeiramente, para a realização desta dissertação foi definida uma metodologia de trabalho. Esta metodologia foi inspirada no processo *CRISP-DM* (*Cross-Industry Standard Process for Data Mining*). O *CRISP-DM* é um modelo de processo padrão aberto que descreve abordagens comuns usadas por especialistas em *data mining* [18]. Na figura 3.1 está presente um diagrama que apresenta quais as etapas deste processo e como se interligam.

Como se pode ver pela figura 3.1 existem seis etapas. A descrição destas etapas é a seguinte:

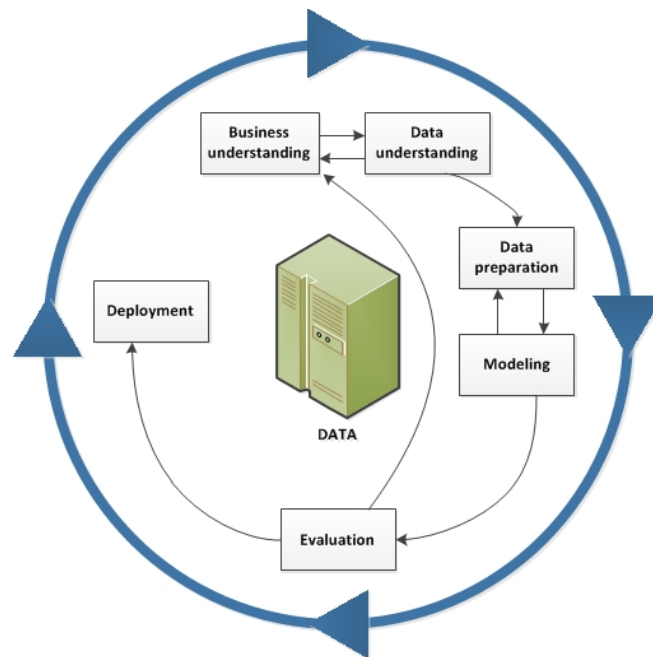


Figura 3.1: Diagrama de fluxo do processo *CRISP-DM* [18].

- **Entendimento do negócio**

Esta é a primeira tarefa do processo. Esta tarefa serve para entender qual é o propósito da análise e fornecer uma definição clara e precisa do problema. Aqui é onde, idealmente, se falaria com os parceiros de negócio, ou interessados, e se perguntava quais eram as questões a ser resolvidas com este projeto. Portanto, com o final desta tarefa deverá surgir um conjunto de questões restritas, as questões principais.

- **Entendimento dos dados**

Esta etapa é a etapa onde se olha para os dados e, normalmente, envolve as seguintes tarefas:

1. Determinação de quais são os dados necessários;
2. Recolha dados, caso os dados necessários não estejam disponíveis;
3. Exploração de dados;
4. Verificação da qualidade dos dados.

- **Preparação dos dados**

Esta etapa é onde ocorre o processamento dos dados com que se irá trabalhar. Pode ser dividido nas seguintes tarefas:

1. Limpeza dos dados;
2. Criação de novos atributos necessários para análise;
3. Integração de dados.

- **Modelação**

Nesta etapa existem tarefas como:

1. Seleção da técnica de modelação;
2. Testagem da qualidade de validade do modelo;
3. Construção do modelo (configurações de parâmetros);
4. Classificação dos modelos.

Esta é a etapa onde são encontradas as respostas às perguntas principais. Estas são encontradas, normalmente, realizando tarefas de modelação e análise estatística.

- **Avaliação**

A etapa de avaliação é onde se avalia o modelo criado. É nesta etapa que se verifica se o modelo criado atinge os objetivos definidos na primeira fase do processo.

- **Desenvolvimento**

Esta é a última etapa do processo. Nesta etapa são feitos esforços de organizar e disponibilizar o que resultou do modelo aplicado aos clientes/interessados.

A metodologia *CRISP-DM* não foi seguida na totalidade, pois o objetivo desta dissertação é a criação de ferramentas que possam ajudar os bombeiros do [RSB](#) a conhecer, analisar e detetar padrões nos seus dados das ocorrências. Portanto, alguns passos foram adaptados para o contexto deste trabalho, nomeadamente, a etapa de modelação onde irão existir tarefas como a de *clustering* e de criação de *dashboards*.

Seguindo o processo escolhido, tudo começa com o entendimento do negócio, neste caso, com o entendimento do que é pretendido do [RSB](#) com o desenvolvimento deste trabalho. Inicialmente houve a oportunidade de ser realizada uma reunião com a [CML](#) onde foram discutidas algumas questões sobre o propósito do trabalho. Após esta reunião foi feito um levantamento das questões principais a que este trabalho teria de responder. A lista das questões principais a serem respondidas com este trabalho é a seguinte:

- Quais são as áreas de influência dos quartéis que apresentam comportamentos semelhantes/diferentes em termos de ocorrências e suas características?
- Quais são as freguesias que apresentam comportamentos semelhantes/diferentes em termos de ocorrências e suas características?
- Existem *hotspots* de ocorrências de certas tipologias no concelho de Lisboa? Esses *hotspots* movem-se ao longo do tempo?
- Existem tipologias de ocorrências que têm o mesmo comportamento temporal?
- Existem agrupamentos espaciais de ocorrências no concelho de Lisboa?

- O que acontece em termos de ocorrências em redor de pontos de interesse do concelho de Lisboa?
- O estado da meteorologia tem impacto no registo de ocorrências no concelho de Lisboa?

Com a lista de questões principais definida segue a fase de entender os dados com os quais se irá trabalhar. Nesta fase foram definidos quais os conjuntos de dados que seriam necessários para o desenvolvimento desta dissertação. Inicialmente já tinham sido disponibilizados pela CML dois conjuntos de dados, o das ocorrências do RSB e dados meteorológicos, no entanto, também seriam necessários dados sobre as áreas de influência dos quartéis que foram solicitados. Depois da definição de quais os dados a utilizar e da sua recolha, foi feita uma exploração e verificação dos mesmos. A exploração/familiarização com os dados foi feita para todos os conjuntos de dados, no entanto, sendo o conjunto de dados das ocorrências o principal, este foi alvo de uma análise relativamente à sua qualidade. Desta análise resultou um relatório de qualidade de dados que pode ser consultado no apêndice A.

Após definidos quais os conjuntos de dados a ser utilizados e, após a sua exploração e verificação de qualidade, houve necessidade de fazer uma preparação destes dados, ou seja, um processamento. Este processamento tem como objetivo preparar os dados para posterior análise e para aumentar a sua qualidade. Os conjuntos de dados utilizados serão apresentados na seção 3.3 assim como o processamento de que foram alvo.

Com os dados prontos para a análise, seguiu-se a fase de modelação. Esta fase, nesta dissertação, consiste na modelação de visualizações, de *dashboards* e interações possíveis. Esta modelação, inicialmente, era feita através da criação de *sketches* em papel. Estes permitiam desenhar o *dashboard*, o conjunto de visualizações, e as interações presentes nele, as ações disponíveis e quais seriam as visualizações afetadas por elas. Nestes *sketches* também constavam as questões mais específicas a que as diferentes visualizações pretendiam responder com o intuito de, com todo o conjunto de visualizações, responder à questão principal que o *dashboard* tencionava resolver. Este processo estimula primeiro o pensar e só depois o fazer e não o contrário. Fazer primeiro sem pensar no assunto não é uma solução válida, pois desenvolver o conjunto de visualizações e interações primeiro e só depois pensar se aquele trabalho desenvolvido responde às questões, por um lado, pode ser uma perda de tempo porque a resposta pode ser que não responde às questões, mas também porque podem existir propostas melhores para a resposta às perguntas. Foi nesta fase de modelação que se sentiu a necessidade de criar uma nova linguagem gráfica para efeitos de auxílio no processo de *design* de *dashboards* e para ajudar a documentar o *dashboard*, nomeadamente, quais as visualizações e interações existentes nele. A criação desta linguagem, denominada "IVML- Interaction Visualization Modeling Language", está sendo desenvolvida em conjunto com outros colegas cujas dissertações são também no âmbito da área de visualização de dados. A criação desta linguagem gráfica não é o objetivo desta dissertação e, por isso, está sendo desenvolvida à parte. No entanto, o IVML será utilizado



para descrever os vários *dashboards* desenvolvidos nesta dissertação. A descrição, através de diagramas IVML, do trabalho desenvolvido está presente na seção 3.5. Depois de estarem definidas quais as visualizações e interações presentes no *dashboard* está na altura de passar do papel para o *Tableau Desktop*, ferramenta utilizada para o desenvolvimento dos *dashboards*.

Depois das ferramentas estarem desenvolvidas, seguiu-se a fase de avaliação das mesmas. Durante esta fase será verificado se os utilizadores, olhando para as ferramentas desenvolvidas, conseguem responder de forma correcta a um conjunto de questões que lhes vai sendo apresentado. Nesta avaliação também será retirado o tempo de resposta dos utilizadores.

Em último surgiu a fase de desenvolvimento, isto é, organizar tudo o que foi desenvolvido e disponibilizar aos interessados as ferramentas. Como as ferramentas foram desenvolvidas utilizando um *software* que necessita de estar instalado e de uma licença paga houve a necessidade de contornar esta questão. Para a apresentação das ferramentas aos interessados não seria desejável forçar a obtenção de *software* pago, por isso as ferramentas foram disponibilizadas no *Tableau Online*, mas não de forma pública, visto estar a trabalhar com dados que não são do domínio público e tem de haver uma preocupação com a proteção dos mesmos. Desta forma não existe a necessidade de instalação de *software* nem da obtenção de licenças. Com a estrutura de acesso às ferramentas já definida, houve também a preocupação de fornecer aos interessados informações de como estas funcionam. Foram disponibilizados vídeos, tipo tutoriais, que exemplificam o uso desses *dashboards*. Futuramente, se o [RSB](#) desejar utilizar as ferramentas desenvolvidas não fica obrigado a obter licenças *Tableau*. Os *dashboards* e visualizações podem ser implementados com linguagens de programação sem necessidade de custos adicionais.

## 3.2 Tecnologias Utilizadas

Actualmente existem várias tecnologias que ajudam no desenvolvimento de projetos. Algumas tecnologias ajudam no processo de manipulação e processamento de dados, normalmente de grande volume, e outras permitem a criação de visualizações de forma fácil e intuitiva onde, muitas das vezes, nem é necessário recorrer à escrita de código para gerar as visualizações e interações com elas. A utilização destas tecnologias foram fundamentais para o desenvolvimento desta dissertação.

Nesta seção será descrito qual o conjunto de *software* que foi utilizado para a realização do trabalho e o porquê de terem sido escolhidos. Nesta lista de *softwares* utilizados constam: o *Tableau Desktop* e *Tableau Online* para o desenvolvimento de visualizações, o *Tableau Prep* e QGIS para a manipulação de dados, o Anaconda para o desenvolvimento de código *Python* para o desenvolvimento de programas relacionados com a temática de *clustering* e, por fim, o *Draw.io* para a criação de diagramas gráficos.

### 3.2.1 Tableau Desktop

O *Tableau Desktop* é um software de criação de visualizações e *dashboards* altamente interativos. Este *software* não é grátis, tem um custo de 70\$ por mês por utilizador, no entanto, disponibiliza licenças grátis para estudantes. É possível se conectar a dados localmente ou na nuvem e não existe necessidade recorrer à escrita de código para aceder e combinar dados díspares. Existem várias outras funcionalidades como a criação rápida de cálculos a partir de dados existentes, criação de linhas de referência, de previsões, e resumos estatísticos. Apesar de o *Tableau Desktop* não ser um Sistema de Informação Geográfica (SIG) este consegue ter muitas funcionalidades geográficas, é capaz de mapear coordenadas geográficas no mapa, consegue-se conectar a vários ficheiros geográficos como *Shapefile* e *GeoJSON*, permite que áreas administrativas, códigos postais e regiões sejam mapeados rapidamente, mas também é possível utilizar geocodificação personalizada. O *Tableau Desktop* tem integração nativa com o provedor de mapas *Mapbox* e existe a possibilidade de utilizar extensões para estender as funcionalidade deste *software*. Por fim, o *Tableau Desktop* tem a funcionalidade de criar *dashboards* com as visualizações desenvolvidas e possibilita a criação de interações entre as visualizações presentes nele como ações de destaque e/ou filtragem. Também é possível criar histórias, isto é, *slides* que deixam de ser estáticos e passam a ser interativos [36].

Relativamente o modo de utilização do *Tableau Desktop* para a criação de visualizações, primeiramente, é necessário conectar-se a um, ou mais, conjunto de dados e verificar se os tipos de dados estão corretos, e, caso sejam dados geográficos, atribuir um significado, por exemplo, se são coordenadas geográficas. O *Tableau Desktop*, de seguida, irá dividir os dados em dimensões e medidas. As dimensões são informações que dizem respeito a valores qualitativos, que estão normalmente sob a forma de *String* e podem ser, por exemplo, datas, categorias ou nomes. As medidas são informações que dizem respeito a valores quantitativos, isto é, informações que podem ser medidas ou agregadas, por exemplo, o valor de vendas pode ser agregado para formar uma média. Depois de tudo estar formatado, é possível criar fórmulas a partir das dimensões e/ou medidas, ou seja, criar novos campos calculados. Agora é possível passar para a criação das visualizações. O *Tableau Desktop* tem uma funcionalidade de *drag and drop*, podemos arrastar para as linhas e para as colunas do gráfico as dimensões e/ou medidas desejadas, podemos também arrastar dimensões e/ou medidas para mapear para variáveis visuais como, por exemplo, cor, tamanho e forma. Neste ponto também é possível adicionar filtros e definir o que aparece na *tooltip* da visualização em construção. À medida que a visualização vai sendo desenvolvida existe um painel que apresenta a pré-visualização do gráfico gerado. De salientar que o *Tableau Desktop* apresenta sugestões de qual o tipo de gráfico que melhor se adequa às dimensões/medidas selecionadas. Sobre o desenvolvimento de *dashboards* no *Tableau Desktop*, é necessário ter as visualizações que irão para o *dashboard* atempadamente concluídas. Para a construção de *dashboards* basta arrastar as visualizações para o sítio pretendido e, caso seja necessário, arranjar o formato dessas visualizações. Por

fim, pode-se criar ações nos *dashboards* para permitir uma maior interatividade, para isso basta escolher qual o tipo de ação pretendia e qual a visualização origem da ação e quais as visualizações que irão ser afetadas.

O *Tableau Desktop* foi utilizado nesta dissertação para a criação de visualizações interativas, de *dashboards* e de histórias. A escolha deste *software* recai na existência de uma licença de estudante com duração de um ano, que pode ir sendo renovada e que dá acesso a todas as ferramentas de desenvolvimento do *Tableau*. Também foi escolhida porque não existe a necessidade de escrever código e é possível produzir visualizações e *dashboards* altamente interativos o que é bastante importante para a exploração e análise de dados. Por outro lado, a tecnologia funciona com *drag and drop* o que a torna bastante poderosa e interessante de utilizar.

### 3.2.2 Tableau Online

O *Tableau Online* é um produto da mesma empresa que desenvolveu o *Tableau Desktop*. É uma plataforma de análise que está hospedada totalmente na nuvem. Podemos dizer que o *Tableau Online* é um *Tableau Desktop*, mas que corre em qualquer *browser*, não havendo assim necessidade de instalar *software* no computador. No entanto, o *Tableau Online* apresenta algumas limitações em termos de funcionalidades relativamente à versão *Desktop*, nomeadamente a criação de alguns tipos de interações com visualizações. Porém é uma ferramenta que é fácil de aceder, via *browser* ou aplicação móvel, e contém funcionalidades como: criar e publicar *dashboards* que podem ser partilhados com qualquer pessoa. Permite a possibilidade de convidar clientes ou interessados para explorar e comentar os dados através de visualizações que podem ser altamente interativas [37].

O *Tableau Online* foi utilizado nesta dissertação como um meio para receber comentários sobre o trabalho que ia sendo desenvolvido. A escolha desta tecnologia recai sobre, primeiramente, possibilitar a publicação *online* dos *dashboards* desenvolvidos (através do *Tableau Desktop*) e por permitir a partilha dos mesmos com várias entidades, neste caso, as entidades interessadas, a Câmara Municipal de Lisboa e o Regimento de Sapadores Bombeiros. Desta forma, os interessados podem aceder ao conteúdo de uma forma muito fácil, através de um *browser*. Para tal, foram criados perfis para as entidades interessadas e foram dadas permissões para aceder, explorar e comentar os *dashboards* submetidos para avaliação. Deste modo é possível haver interação com as ferramentas disponibilizadas de pessoas que não participaram no processo de criação das mesmas e receber *feedback* pertinente sobre a qualidade do trabalho desenvolvido. Portanto, de modo a obter comentários e dar a conhecer algumas das ferramentas desenvolvidas, foi necessário recorrer a esta tecnologia, o *Tableau TableauOnline*.

### 3.2.3 Tableau Prep

O *Tableau Prep* é outro *software* da empresa que desenvolveu o *Tableau Desktop* e é composto por dois produtos, o *Prep Builder*, que permite combinar, modelar e limpar

dados para análise, e o *Prep Conductor*, que permite partilhar fluxos de dados e gerir os mesmos. Esta tecnologia permite fazer a preparação e processamento de dados de forma bastante intuitiva e fácil, sem necessidade de recorrer à escrita de código. O *Tableau Prep* torna mais intuitiva a combinação, limpeza, modelação e partilha dos dados, pois oferece uma *interface* visual que permite ter uma imagem completa dos dados e recursos inteligentes para tornar a limpeza, automação e administração mais fácil. Entre as várias funcionalidades há que destacar: a união e junção de conjunto de dados, transformação de linhas em colunas e vice-versa, agregações, filtragens e mudança do tipo de dados. O *software* ajuda o desenvolvimento de uma solução completa de preparação de dados baseada em fluxos, isto é, possibilita a existência de uma cadeia de passos, onde cada passo será uma alteração, processamento, nos dados, por exemplo, a remoção ou adição de uma coluna, e no final é possível guardar o resultado desse fluxo, o novo conjunto de dados processado, em vários tipos de ficheiros [38].

O *Tableau Prep* foi utilizado nesta dissertação para efectuar algum processamento de dados, nomeadamente, nos dados das ocorrências que necessitavam de algumas intervenções. A escolha deste *software* recai na possibilidade de integração com o *Tableau Desktop*, no sentido em que é possível abrir a saída do *Prep Builder* (os dados que estão a ser alvo de processamento) no *Tableau Desktop*, facilitando a permanência no fluxo de análise, não existindo, assim, a necessidade de sair do fluxo para efectuar novas alterações aos dados. Outro ponto a favor para a decisão da utilização desta tecnologia foi o facto de esta conter uma *interface* visual onde não é necessário recorrer à escrita de código para efectuar as alterações aos dados.

### 3.2.4 QGIS

O QGIS é um *software* com código aberto de Sistema de Informação Geográfica (SIG). Esta ferramenta é disponibilizada gratuitamente e funciona em diferentes sistemas operativos como Linux, Unix, Mac OSX, Windows e Android. O QGIS possibilita a criação, gestão, visualização, manipulação e análise de dados georreferenciados. Esta ferramenta permite a criação de mapas com múltiplas camadas que podem usar diferentes projeções. As camadas podem ser geradas a partir de diferentes tipos de ficheiros como, por exemplo, *csv*, *raster*, *shapefiles* ou *GeoJSON*. Os dados podem ser considerados/guardados como sendo pontos, linhas ou polígonos. O QGIS tem integração de uma consola *Python*, que pode ser utilizada para desenvolver novas funcionalidades como algoritmos para manipular dados utilizando código na linguagem *Python*. O QGIS disponibiliza também *plugins* que vão sendo adicionados pela comunidade de código aberto, aumentando, assim, o número de funcionalidades desta ferramenta [32].

O QGIS foi utilizado nesta dissertação para efetuar algum processamento de dados e para criar novos conjuntos de dados a partir dos dados das ocorrências, nomeadamente, conjuntos de dados que contêm informações sobre quais os pontos que pertencem a um polígono. A escolha desta tecnologia recai no facto de esta ser gratuita e apresentar todas

as funcionalidades que são necessárias para este trabalho, nomeadamente, a inserção de dados georreferenciados no mapa e algoritmos que permitem detetar quais são os pontos que estão inseridos nos polígonos.

### 3.2.5 Anaconda

O Anaconda é uma distribuição gratuita e de código aberto das linguagens de programação R e *Python* para computação científica que visa simplificar a gestão de pacotes e implantação. O Anaconda é abrangentemente utilizado para a ciência de dados, processamento de dados em grande escala, aprendizagem automática e análise preditiva. A distribuição está disponível para diferentes sistemas operativos, Windows, Linux e macOS. O Anaconda disponibiliza algumas aplicações como *JupyterLab*, *Jupyter Notebook*, *QtConsole*, *Spyder*, *Glue*, *Orange*, *RStudio* e *Visual Studio Code* [19].

Nesta dissertação a aplicação com mais utilização é o *Spyder* (*Scientific Python Development Environment*). O *Spyder* é um ambiente de desenvolvimento integrado (IDE) gratuito que está incluído com o Anaconda. Como IDE é composto por um editor, uma consola *IPython*, um explorador de variáveis, um *debugger* e um menu de ajuda. É um poderoso ambiente científico escrito em *Python*, para *Python* e projetado para cientistas, engenheiros e analistas de dados. Apresenta uma combinação de funcionalidades de edição avançada, análise, *debugging* e funcionalidade de criação de perfil de uma ferramenta de desenvolvimento abrangente com a exploração de dados, execução interativa, inspeção profunda e capacidades de visualização de um pacote científico. O *Spyder* integra-se a vários pacotes proeminentes do *Python*, o *NumPy*, *SciPy*, *Matplotlib*, *pandas*, *scikit-learn* entre outros. É possível estender mais as funcionalidades desta tecnologia com *plugins* [10].

O Anaconda, mais especificamente o IDE *Spyder*, foi utilizado nesta dissertação para processamento de dados e para o desenvolvimento de programas em *Python* que fazem uso de algoritmos de *clustering*. A escolha desta tecnologia recai no facto de a linguagem *Python* ser versátil e fácil de usar e porque, para esta dissertação, contém muitas bibliotecas interessantes, nomeadamente, bibliotecas de aprendizagem automática. A escolha do Anaconda e do *Spyder* recai também no facto de todo o *set up* estar já montado.

### 3.2.6 Draw.io

O *Draw.io* é um *software* de desenvolvimento de diagramas de código aberto. Está disponível *online*, via *browser*, e existem também uma versão *desktop*. Esta tecnologia disponibiliza de raiz diferentes tipos de diagramas como diagramas de classe, fluxogramas, diagramas de entidade-relação e diagramas de sequência. O *Draw.io* permite a criação de vários tipos de diagramas e possibilita o seu armazenamento em diferentes tipos de ficheiros, nomeadamente, PNG, JPEG, SVG, PDF, XML e HTML. Também permite guardar o diagrama num formato especial que pode ser recarregado novamente no *browser* para editar. Existe também a possibilidade de criação de bibliotecas com diferentes formas

pré-definidas. Por fim, o *Draw.io* tem integração com várias plataformas como *Google Drive*, *OneDrive*, *Github*, *GitLab* e *Dropbox*, existindo também uma aplicação para o *Word*, *Powerpoint* e *Excel* do *Office 365* [13].

O *Draw.io* foi utilizado nesta dissertação para a criação de diagramas e para a criação de uma biblioteca gráfica com várias formas. A escolha desta tecnologia recai no facto de ser gratuita e ser *online*, acessível via *browser*, e por permitir o armazenamento de várias formas em formato de uma biblioteca que pode ser descarregada e recarregada posteriormente. Também permite descarregar o diagrama criado num ficheiro editável que pode ser recarregado novamente no *Draw.io* e fazer novas alterações.

### 3.3 Processamento de Dados

Para a realização desta dissertação é crucial conhecer e entender os dados com os quais se vai trabalhar. No entanto, também é muito importante que os dados tenham qualidade para se poder trabalhar com eles.

Nesta seção irão ser apresentados os dados que foram utilizados nesta dissertação, nomeadamente, os dados das ocorrências do *RSB* e dados de meteorologia do *IPMA*, também será apresentado o processamento feito a estes conjuntos de dados e respetivos motivos.

#### 3.3.1 Dados das Ocorrências do RSB

Os dados principais necessários para a realização desta dissertação foram disponibilizados pela Câmara Municipal de Lisboa. Estes dados são relativos às ocorrências do Regimento de Sapadores Bombeiros que aconteceram entre os anos de 2011 e 2018.

No geral, cada ocorrência contém informações sobre a sua localização, dados temporais e características.

A Tabela 3.1 apresenta todos os atributos das ocorrências que estão relacionados com a sua localização. Como podemos observar na Tabela 3.1, existem nove atributos que são

Tabela 3.1: Informação de localização das ocorrências

Atributo	Descrição
Longitude	Longitude do local da ocorrência
Latitude	Latitude do local da ocorrência
Código de via	Código da via, isto é, da rua do local da ocorrência
Código SIG	Código SIG, por exemplo, do edifício ou, novamente, código da via do local da ocorrência
Código SIG da fracção	Código SIG da fracção do local da ocorrência
Complemento do local	Detalhes sobre o local da ocorrência
Designação do local	Morada do local da ocorrência
Código SIG da freguesia	Código SIG da freguesia do local da ocorrência
Descrição da freguesia	Nome da freguesia do local da ocorrência

relativos à localização da ocorrência. Dois desses atributos são coordenadas geográficas, latitude e longitude. Quatro atributos são códigos que codificam o local da ocorrência: o código da via; o código SIG; o código da fracção e o código SIG da freguesia. Os restantes três atributos, relativos à localização das ocorrências, são atributos simbólicos: o complemento do local; a designação do local e a descrição da freguesia.

O atributo do código SIG tem algumas particularidades, pois pode conter também o código da via. Esta situação surge em ocorrências que aconteceram na rua e não, por exemplo, em casas.

Visto que os dados são referentes aos anos de 2011 até 2018, os códigos SIG e descrição das freguesias de anos posteriores a 2013 podem conter informações diferentes dos anos anteriores a 2013. Isto acontece porque no ano de 2012 houve uma reorganização administrativa que levou à união de várias freguesias do concelho de Lisboa.

A Tabela 3.2 exibe todos os atributos temporais das ocorrências. Como a Tabela 3.2 demonstra, apenas existem dois atributos temporais. Um refere-se à data e hora do registo da ocorrência no sistema. O outro é relativo à data e hora da última mudança de estado que a ocorrência sofreu.

Tabela 3.2: Informação temporal das ocorrências.

Atributo	Descrição
Data e hora	Data e hora do registo da ocorrência
Data e hora alteração do estado	Data e hora da alteração do estado da ocorrência

A Tabela 3.3 contém todos os atributos relativos às características das ocorrências. É possível verificar pela Tabela 3.3 que existem quatorze atributos que caracterizam as ocorrências. Cinco destes atributos são identificadores: o identificador da ocorrência; do estado; da tipologia; da entidade que iniciou o processo e o identificador do sistema do RSB. Quatro atributos são simbólicos: a descrição do estado; a tipologia da ocorrência e a entidade que iniciou o processo. Três atributos são numéricos: o indicador do grau de risco; o número de viaturas a enviar para o local e a estimativa de elementos a enviar também para o local da ocorrência. Por fim, os restantes dois atributos são booleanos: o indicador de suspensão e de falso alarme após a ativação de meios para o local da ocorrência.

A descrição do estado da ocorrência apenas pode apresentar um de três valores possíveis. Os três valores possíveis e respectivos significados são os seguintes:



Tabela 3.3: Informação das características das ocorrências.

Atributo	Descrição
ID	Identificador único da ocorrência
ID do estado	Identificador do estado da ocorrência
Descrição do estado	Descrição do estado da ocorrência
ID da tipologia	Identificador da tipologia da ocorrência
Descrição da tipologia	Descrição da tipologia da ocorrência
Descrição da ocorrência	Detalhes sobre a ocorrência
ID da entidade	Identificador da entidade que iniciou a ocorrência
Descrição da entidade	Sigla da entidade que iniciou a ocorrência
Indicador do grau de risco	Grau de risco associado à ocorrência
Indicador de suspensão	Indica se a ocorrência foi suspensa ou não
Identificador de falso alarme	Indica se a ocorrência foi falso alarme depois da ativação de meios ou não
ID sistema de ocorrências do RSB	Código identificador do sistema de ocorrências do RSB
Número de viaturas	Número de viaturas a enviar para o local da ocorrência
Estimativa número de elementos	Estimativa do número de elementos que serão enviados para o local da ocorrência

- "A": acionada.

Este estado significa que a ocorrência tem meios a trabalhar no local da intervenção.

- "FO": fecho operacional.

Este estado significa que já não existem meios no local da intervenção, ou seja, a intervenção acabou.

- "FA": fecho administrativo.

Este estado só pode acontecer após a ocorrência estar fechada operacionalmente. Este estado significa que a ocorrência está fechada operacionalmente e que já foi realizado um relatório relativo à intervenção.

As ocorrências têm uma tipologia associada. Esta tipologia pode ser separada em duas partes, um tipo geral e um subtipo. Nestes dados foram detectas várias tipologias gerais que estão presentes na tabela 3.4.

Existem algumas tipologias gerais que são mais comuns nos dados das ocorrências que outras. A figura 3.2 apresenta a distribuição das ocorrências pelas tipologias gerais. Como ilustrado na figura 3.2, quase metade das ocorrências presentes neste conjunto de dados, cerca de 49,5%, é referente a ocorrências do tipo "Serviços", sendo que existe uma grande diferença relativamente à segunda tipologia geral que mais ocorrências tem, a de "Infra-estruturas e Vias de Comunicação" que contém cerca de 13,5% das ocorrências do RSB. Para maior compreensão das tipologias gerais a figura 3.3 apresenta a evolução do número de ocorrências das tipologias gerais ao longo das semanas de sete anos de dados.



Tabela 3.4: Tipologias gerais presentes no conjunto de dados das ocorrências.

Tipologia Geral
Acidentes
Actividades
Conflitos Legais
Eventos de Protecção Civil
Incêndio
Infra-estruturas e Vias de Comunicação
Pré-Hospitalar
Serviços
Tecnológicos Industriais

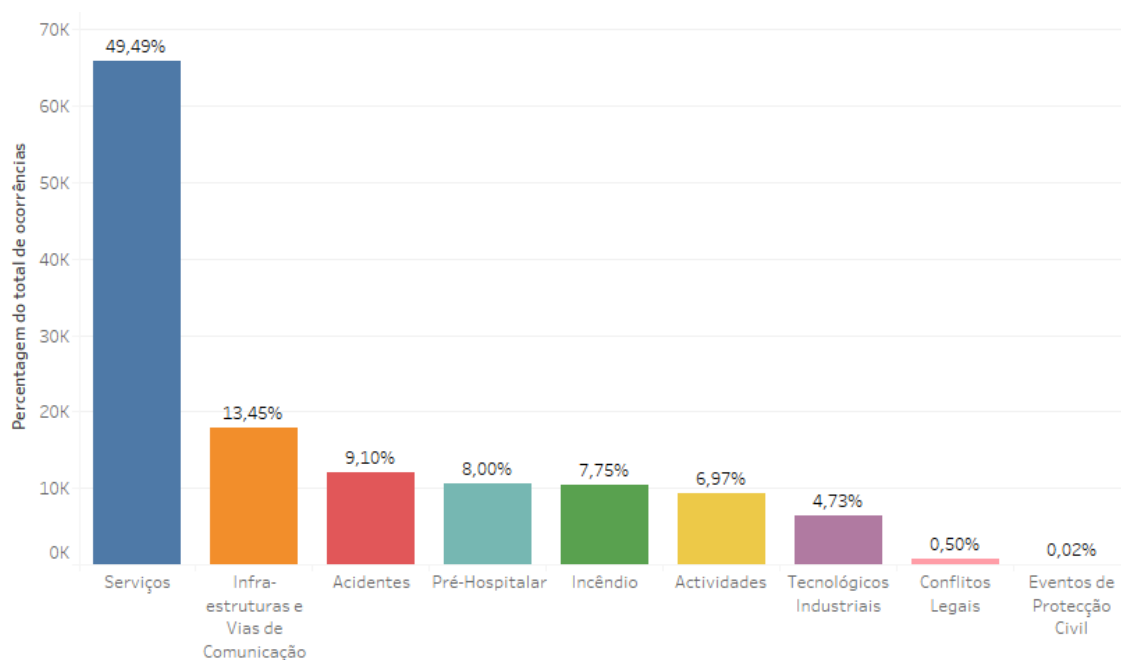


Figura 3.2: Distribuição do número de ocorrências pela tipologia geral.

De salientar que esta figura tem um eixo vertical com valores máximos diferentes para tipologia geral.

Com a figura 3.3 é possível detetar alguma sazonalidade nas tipologias gerais. Claramente existe sazonalidade nos incêndios, existe um aumento no número de ocorrências que é periódico e acontece durante os meses tipicamente mais quentes. É também possível analisar que a tipologia de infra-estruturas e vias de comunicação apresenta picos elevados no número de ocorrências. Estes picos poderão ser um interesse de futura análise e poderão estar relacionados com eventos meteorológicos mais severos, como inundações. Também se pode observar que o número de ocorrências da tipologia geral de "Actividades" tem vindo a ter um aumento ao longo do tempo, assim como a tipologia geral de "Acidentes". Tipologias do tipo "Eventos de Protecção Civil" apenas começaram a surgir no ano de 2017. De resto, as restantes tipologias gerais não apresentam nenhum padrão

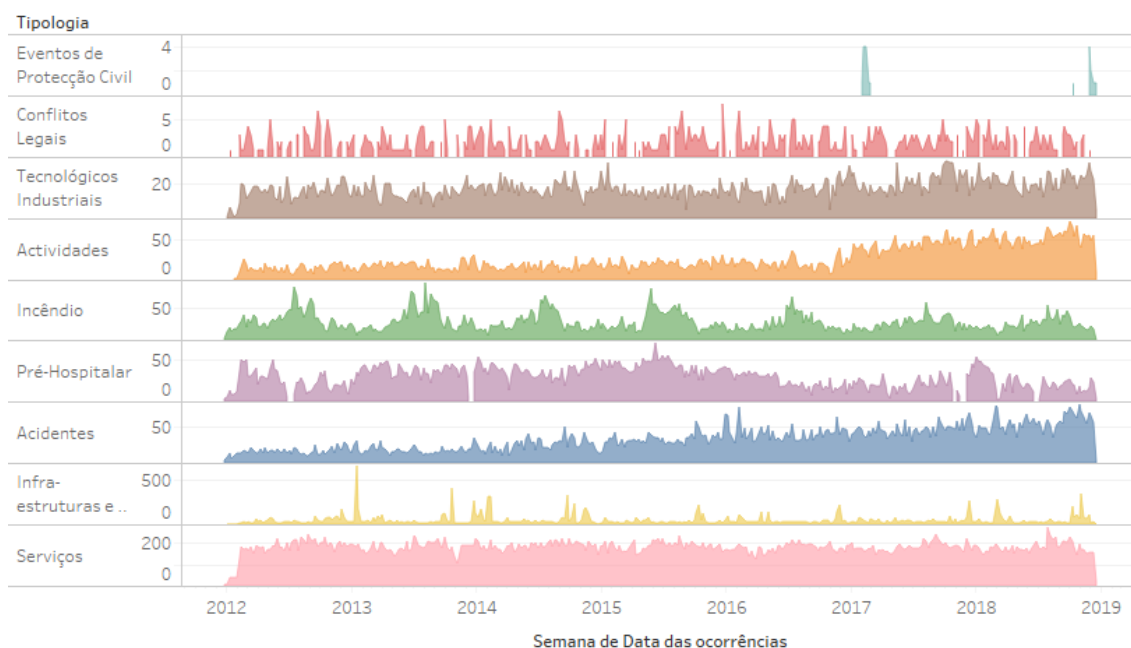


Figura 3.3: Evolução temporal das tipologias gerais das ocorrências ao longo de todas as semanas existentes no conjunto de dados.

temporal visível.

Cada tipologia geral tem associado um conjunto de subtipos possíveis. Por exemplo, para ocorrências do tipo geral "Pré-Hospitalar" existem seis subtipos: "Afogamento", "Doença Súbita", "Intoxicação", "Parto", "Queimado" e "Traumatismo/Queda". Uma lista com todos os subtipos de todas as tipologias gerais existentes está presente no apêndice B.

Para uma determinada tipologia de ocorrência existe uma referência da quantidade de elementos e viaturas a enviar para o local da ocorrência. As viaturas só podem sair para o local da intervenção se tiverem a guarnição completa.

O atributo do grau de risco apresenta dois valores numéricos possíveis. Os valores possíveis para este atributo e respectivos significados são:

- 2: grau de risco elevado;
- 1: grau de risco moderado.

O atributo da descrição da ocorrência é um campo de texto livre que contém detalhes sobre a ocorrência como, por exemplo, pontos de referência para ajudar a localizar o local. Também é comum aparecer neste atributo, em casos de ocorrências pré-hospitalares, o sexo e idade do paciente e os códigos dos [Centros de Orientação de Doentes Urgentes \(CODU\)](#).

Para explorar e conhecer o conjunto de dados das ocorrências foi elaborado um relatório de qualidade destes dados, não só com o intuito de explorar os dados, mas também de saber qual a qualidade dos mesmos visto serem a base do trabalho desta dissertação.

O relatório de qualidade de dados elaborado pode ser consultado no apêndice [A](#). Este relatório levanta alguns problemas com a qualidade deste conjunto de dados, portanto foi necessário realizar um pré-processamento antes de avançar para o desenvolvimento de ferramentas de exploração e análise.

Foi tomada a decisão de não utilizar, para efeitos de análise, atributos que apresentem percentagens altas de valores nulos apesar de poderem ser atributos interessantes para análise. Este é o caso dos atributos de grau de risco da ocorrência e do indicador de falso alarme após a activação de meios para o local da intervenção, estes apresentam 97% e 99% de valores nulos, respectivamente. Houve a possibilidade de comunicar com a [CML](#) sobre o porquê destes valores serem tão elevados e, relativamente ao grau de risco, foi comunicado que a maior parte das ocorrências não está tipificada, sendo difícil atribuir um valor de grau de risco às ocorrências e, relativamente, ao indicador de falso alarme foi dito que se devia assumir que quando este atributo aparece a nulo é porque o campo não foi preenchido, que apenas é colocado 0 (não foi falso alarme) ou 1 (foi falso alarme) quando se acha necessário ou quando se conseguiu determinar esse facto. Com todas as respostas obtidas não houve dúvida que estes atributos não seriam considerados para análise.

O primeiro processamento passou por excluir todas as ocorrências do ano de 2011. Isto porque este ano não é considerado um ano representativo da realidade, pois apresenta poucos registos de ocorrências e existem alguns meses onde nem há registos.

Existia o problema da coexistência nos dados de duas organizações administrativas diferentes das freguesias. Portanto, houve a necessidade de eliminar uma dessas organizações, a organização mais antiga, e atualizar as ocorrências anteriores ao ano de 2013 (inclusive) para a nova organização. Como houve uma junção das freguesias as alterações não foram complexas de realizar, a única excepção foi a antiga freguesia "Santa Maria dos Olivais" que se dividiu em duas novas freguesias. Portanto, nesta situação de divisão, foi necessário recorrer às coordenadas geográficas das ocorrências para actualizar a freguesia. Outro problema, também relacionado com as freguesias, foi o facto de algumas ocorrências apresentarem o identificador da freguesia preenchido, mas não tinham preenchido a designação da freguesia. Sendo que um identificador mapeia para uma única designação, a designação foi preenchida de acordo com o identificador presente.

Outro processamento que foi executado foi mapear um identificador de tipologia da ocorrência para uma única descrição de tipologia. Acontece que existiam identificadores que apresentavam duas descrições diferentes. Um exemplo deste problema está presente na tabela [3.5](#). Para solucionar esta questão foram identificados quais os identificadores e descrições associadas e foi escolhida apenas uma das descrições possíveis, sendo que as descrições não variam muito entre si. Algumas ocorrências continham também o mesmo problema do que acima descrito para as freguesias, continham o identificador da tipologia, no entanto, não tinham descrição da tipologia associada. Portanto, agora que todos os identificadores mapeiam para uma única descrição, solucionar estes casos foi fácil.

Por fim, existiam alguns problemas com atributos geográficos. O primeiro problema

Tabela 3.5: Identificadores de tipologias a mapear para diferentes descrições de tipologias.

Identificador	Descrições
176	8711 - Actividades - Apoio Social Avaliação 8711 - Actividades - Apoio Social Visitas Pontuais

detetado foi que existiam ocorrências sem coordenadas geográficas, no entanto, tinham uma morada associada. Para ocorrências com este problema foi desenvolvido um programa em *Java* que dada a morada gera as coordenadas geográficas associadas. Para tal foi utilizada a *Geocoding and Search API* do *HERE*. Esta *API REST* é utilizada para encontrar as coordenadas geográficas de um endereço conhecido, local, localidade ou área administrativa, mesmo se a consulta estiver incompleta ou parcialmente incorreta. Também existiam ocorrências que ficavam fora do concelho de Lisboa então foram removidas. Este processo foi efectuado utilizando as funcionalidades do QGIS. Foram criadas duas camadas, uma de pontos, as ocorrências, e outra um polígono, a área do concelho de Lisboa e foram pedidas as ocorrências que estavam dentro desse polígono. Para terminar, para o desenvolvimento de alguns *dashboards* foi necessário criar novas colunas no conjunto de dados. Foi acrescentada uma nova coluna ao conjunto de dados para indicar em que área de influência dos quartéis do *RSB* a ocorrência caiu. Para esta tarefa a *CML* disponibilizou um *GeoJSON* com os polígonos de todas as áreas de influências dos quartéis. Esta situação é semelhante à descrita anteriormente onde foi utilizado o QGIS.

### 3.3.2 Dados de Meteorologia

Para a procura de variáveis potencialmente explicativa para possíveis padrões detetados, a Câmara Municipal de Lisboa cedeu o conjunto de dados meteorológicos, relativos ao ano de 2018, do *IPMA*.

O conjunto de dados meteorológicos disponibilizado é apenas relativo ao ano de 2018 e apresenta as condições meteorológicas. Este conjunto de dados apresenta os atributos que estão presentes na tabela 3.6 e que são recolhidos de hora a hora.

Tabela 3.6: Atributos presentes no conjunto de dados meteorológicos do *IPMA*.

Atributo	Unidades
Rumo do vento	
Intensidade do vento	km/h
Temperatura do ar	°C
Radiação solar	$W/m^2$
Precipitação	mm
Humidade do ar	%

Existindo três estações meteorológicas no concelho de Lisboa, os atributos descritos na tabela 3.6, são apresentados para cada uma dessas estações, de hora em hora. Na figura

3.4 podemos ver onde estão localizadas estas estações meteorológicas. Acontece que nem sempre todas estas estações estão a funcionar ou conseguem recolher as informações dos atributos. Quando isso acontece é apresentado um valor *default* no respectivo atributo para informar que algo não correu bem, esse valor é -990.0.

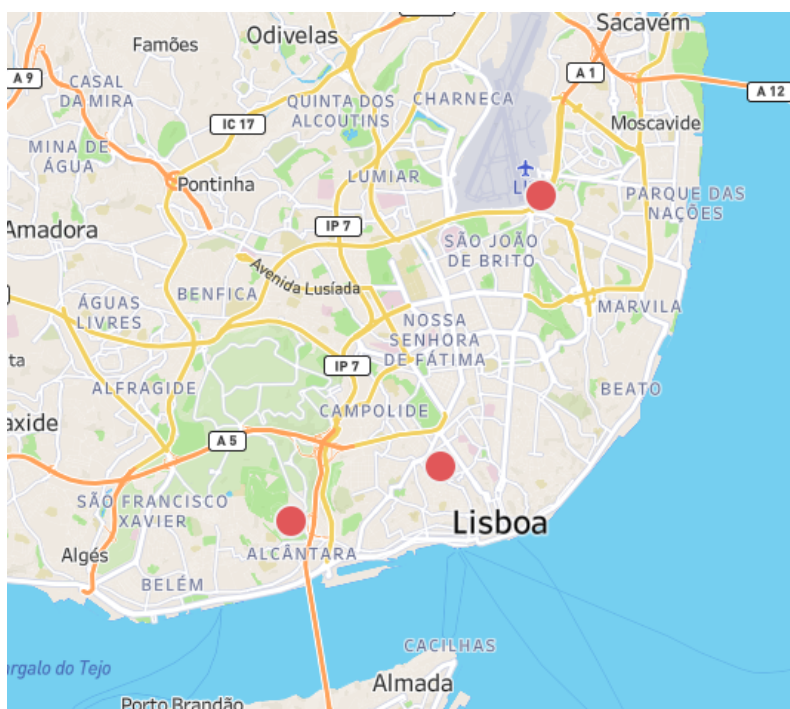


Figura 3.4: Localização das estações meteorológicas automáticas do IPMA em Lisboa.

Não sendo o concelho de Lisboa muito extenso em termos de área e por conter três estações meteorológicas, foi investigado se os valores apresentados por essas estações, para cada atributo, às mesmas horas diferiam muito entre si. Da análise feita resultou que não existe muita diferença entre as estações e, por vezes, apenas uma das estações está a recolher dados. Isto levou à necessidade de processar os dados de meteorologia para ter apenas uma vista global do tempo no concelho de Lisboa, e não três, e também para fugir à falta de dados nalgumas estações.

Foi criado um novo ficheiro que contém as colunas apresentadas na tabela 3.7. O atributo rumo do vento não foi incluído, pois foi considerado desinteressante para as análises que irão ser feitas posteriormente. As colunas "Humidade", "Precipitação", "Radiação", "Temperatura" e "Vento" foram calculadas a partir dos dados iniciais das três estações meteorológicas, sendo a média horária dos valores apresentados por essas estações. Caso não houvesse registo de valores nalguma estação, essa estação não entraria para o cálculo da média. Deste modo ultrapassou-se o facto de algumas estações não estarem a funcionar e conseguiu-se obter uma visão global do estado do tempo no concelho de Lisboa, não deixando de ser representativo, pois inicialmente não existia muita diferença entre os valores apresentados pelas diferentes estações.

Tabela 3.7: Colunas do novo conjunto de dados após processamento.

Coluna	Descrição
Data e hora	Data e hora da leitura
Humidade	Humidade do ar (%)
Precipitação	Precipitação (mm)
Radiação	Radiação solar ( $W/m^2$ )
Temperatura	Temperatura do ar ( $^{\circ}C$ )
Vento	Intensidade do vento (km/h)

### 3.4 Agrupamento

Para a realização deste trabalho, nomeadamente, para conseguir responder a algumas das questões principais foi necessário recorrer a técnicas de *clustering*.

Esta seção está dividida em duas partes, o *clustering* temporal e o espacial. Será descrito o porquê de ser necessário recorrer a estas técnicas de *clustering*, serão apresentados quais os agrupamentos que foram feitos, quais os algoritmos utilizados, como foi feita a escolha dos parâmetros desses algoritmos e os resultados obtidos.

#### 3.4.1 Temporal

Uma das questões principais que foi recolhida no início do trabalho, presente na seção 3.1, foi: "Existem tipologias de ocorrências que têm o mesmo comportamento temporal?". Para responder a esta questão é necessário utilizar técnicas de *clustering* com o objetivo de agrupar as tipologias que têm um comportamento semelhante em termos temporais. O conceito de semelhantes em termos temporais é abrangente, por isso, foi decidido fazer o agrupamento em diferentes granularidades temporais. O agrupamento foi realizado para cada uma destas granularidades temporais:

- Estações do ano;
- Meses;
- Dias da semana;
- Partes do dia - a tabela 3.8 apresenta como foi decomposto um dia por partes;
- Horas.

Definidos quais os agrupamentos que necessários, houve a necessidade de construir os vários conjuntos de dados que servirão para alimentar o algoritmo de *clustering*. A tabela 3.9 exemplifica o conjunto de dados criado para o agrupamento ao nível das estações do ano. Para as restantes granularidades temporais, meses, dias da semana, partes do dia e horas, a construção do conjunto de dados é semelhante, sendo que nas colunas estão dados referentes à partição temporal de cada granularidade temporal. Como demonstrado

Tabela 3.8: Decomposição de um dia por partes.

Parte do dia	Intervalo de tempo
Madrugada	00:00h até 07:59h
Manhã	08:00h até 11:59h
Almoço	12:00h até 14:59h
Tarde	15:00h até 18:59h
Jantar	19:00h até 20:59h
Noite	21:00h até 23:59h

na tabela 3.9, existe uma linha para cada tipologia existente no conjunto de dados e uma coluna para cada partição do tempo. Como existem tipologias que contêm muitas ocorrências e outras que não contêm assim tantas ocorrências não foi utilizado o número de ocorrências absoluto para preencher as linhas do conjunto de dados. Portanto, foi utilizada a percentagem do total de ocorrências de cada tipologia. Isto é, cada tipologia tem um número total de ocorrências associado e para cada partição do tempo são contadas quantas dessas ocorrências aconteceram nessa partição, sendo que, no final esse número de ocorrências na partição é dividido pelo total de ocorrências da tipologia obtendo-se assim um valor que é transformado em percentagem multiplicando por cem. No final, a soma dos valores presentes em cada linha é sempre 100% não havendo assim distinção entre tipologias com muitas e poucas ocorrências. Deste modo é possível analisar apenas o comportamento temporal das várias tipologias esquecendo a sua popularidade.

Tabela 3.9: Exemplo do conjunto de dados criado para o *clustering* das tipologias para as estações do ano.

Tipologia	Inverno	Outono	Primavera	Verão
1100 - Incêndio - Povoamento Florestal	1,8%	10,5%	12,3%	75,4%
4200 - Pré-Hospitalar - Doença Súbita	27,7%	22,5%	27,1%	22,7%

Para o agrupamento temporal, nas várias granularidades, das tipologias foi utilizado o algoritmo de *clustering Kmeans*. A escolha deste algoritmo em particular recaiu no facto de considerar todos os registos para o agrupamento, isto é, não descarta registos, não existe a noção de ruído. A utilização do *Kmeans* proporciona um agrupamento de todas as tipologias em grupos e só existe a necessidade de otimizar um parâmetro, o número de grupos finais, o parâmetro  $k$ . Para a escolha do parâmetro deste algoritmo, o  $k$ , foi utilizado o conhecido método do cotovelo e os valores de *silhouette score*. A figura 3.5 apresenta o gráfico resultante da aplicação do método do cotovelo no conjunto de dados criado para o *clustering* das tipologias para as estações do ano.

Com o gráfico presente na figura 3.5 não é logo imediato a deteção do cotovelo, mas o melhor valor de  $k$  está entre seis e oito. Como não é exato onde se situa o cotovelo, foi utilizado o *silhouette score* para decidir entre os valores seis, sete e oito para o  $k$ . O *silhouette score* mede a semelhança de um ponto com o seu próprio *cluster* (coesão) em



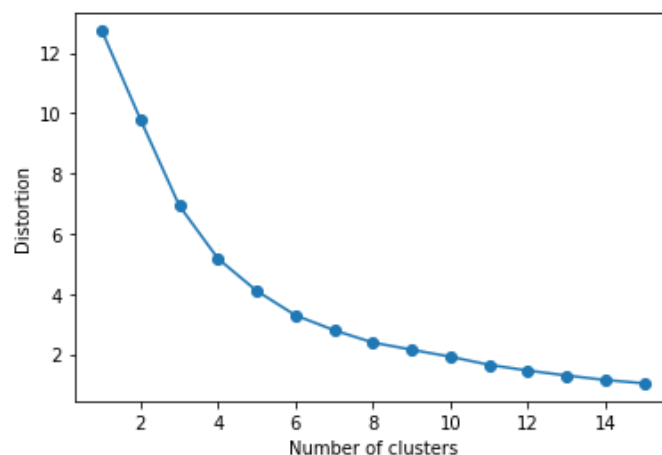


Figura 3.5: Gráfico do método do cotovelo para o conjunto de dados do *clustering* temporal das tipologias para as estações do ano.

comparação com outros *clusters* (separação). O intervalo do valor desta pontuação está entre 1 e -1, sendo que quanto mais alto o valor melhor. Um valor alto é desejável e indica que o ponto está colocado no *cluster* correto. O *silhouette score* atinge o máximo global no  $k$  ideal. Os resultados para o *silhouette score* obtidos para os diferentes  $k$  estão apresentados na tabela 3.10.

Tabela 3.10: Valores da métrica *silhouette score* para diferentes  $k$  no agrupamento por estações do ano.

Valor do $k$	Valor do <i>silhouette score</i>
6	0.463
7	0.436
8	0.425

Com os resultados obtidos presentes na tabela 3.10, o valor do número de *clusters*, o  $k$ , escolhido foi o  $K = 6$ , pois está na zona do cotovelo e tem a melhor pontuação do *silhouette score*. Ao analisar os resultados obtidos com esta definição de  $k$  através de visualizações no *Tableau* notou-se que existiam grupos com poucas tipologias, inclusive, grupos que apenas continham uma tipologia. Devido a casos destes houve a necessidade de intervir, pois visualmente iria afetar a percepção do utilizador. Portanto, foi considerado que grupos com menos de cinco tipologias seriam agrupados. O grupo resultante deste agrupamento seria considerado um grupo de tipologias que não partilham nenhum agrupamento temporal entre si. Para demonstrar que este agrupamento não é significativo, existem no total 109 tipologias e o tamanho do maior grupo que contém os grupos com poucas tipologias é de 16 no caso do agrupamento considerando as horas. Os métodos acima explicados foram utilizados para todas as granularidades temporais.

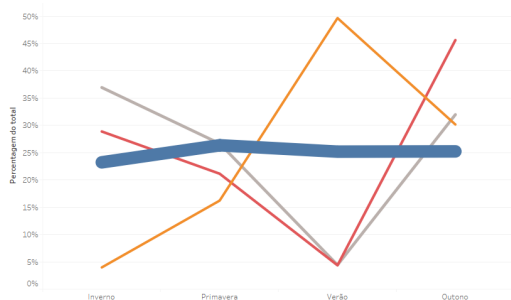
A tabela 3.11 contém uma síntese do agrupamento temporal nas diferentes granularidades temporais consideradas. Esta tabela apresenta o valor escolhido para o  $k$ , o valor



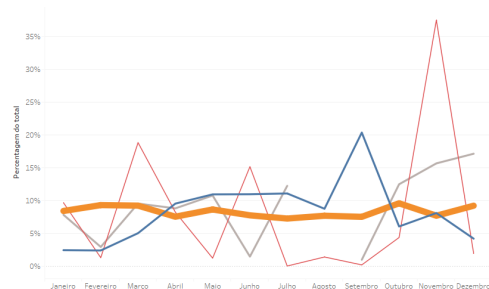
Tabela 3.11: Síntese do agrupamento temporal.

Granularidade	$k$	<i>silhouette score</i>	Nº grupos finais	Nº tipologias por grupos
Estações do ano	6	0.463	4	72; 14; 12 e 11
Meses do ano	12	0.319	4	71; 17; 17 e 4
Dias da semana	10	0.379	4	74; 16; 12 e 7
Partes do dia	6	0.435	4	70; 20; 10 e 9
Horas	10	0.312	4	68; 17; 15 e 9

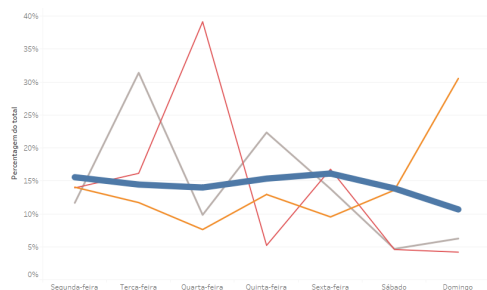
da métrica *silhouette score* associado ao parâmetro escolhido, o número final de grupos temporais resultantes da execução do algoritmo *Kmeans* com o parâmetro escolhido e após a agrupamento dos grupos de pequena dimensão e, por fim, o número de tipologias de cada um desses grupos.



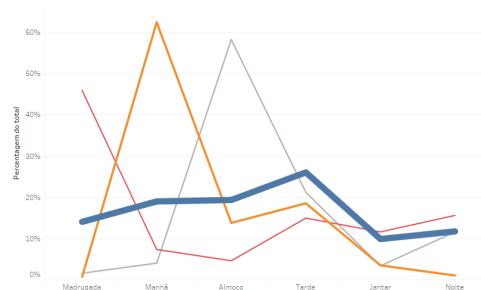
a Resultado para as estações do ano.



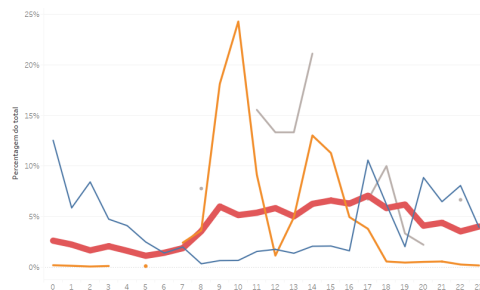
b Resultado para os meses do ano.



c Resultado para os dias da semana.



d Resultado para as partes do dia.



e Resultado para as horas.

Figura 3.6: Resultados do agrupamento temporal.

A figura 3.6 apresenta os resultados obtidos para os vários agrupamentos temporais das tipologias. As figuras presentes na figura 3.6 apresentam o comportamento temporal

dos diferentes grupos resultantes do agrupamento. Cada linha representa o comportamento de um grupo e a grossura da linha está mapeada para o número de tipologias contidas nesse grupo, ou seja, quanto mais grossa a linha maior o número de tipologias esse grupo contém. As linhas representadas a cor cinzenta apresentam o grupo que sofreu a agregação dos grupos que continham poucas tipologias. Em alguns destes gráficos existe uma descontinuidade na linha que representa o comportamento temporal dos grupos. Isto quer dizer que não existem registos de ocorrências daquele grupo naquele dado momento do tempo. A figura 3.6a apresenta o comportamento temporal, ao longo das estações do ano, dos quatro grupos resultantes. O grupo a azul representa tipologias que têm um comportamento temporal constante ao longo das estações. Este grupo é o grupo com mais tipologias associadas. Isto é visível pela grossura da sua linha. O grupo a laranja representa tipologias que têm maior incidência durante a estação de Verão, sendo que tem pouca expressividade durante o Inverno. O grupo a vermelho representa tipologias que têm muita incidência no Outono e alguma no Inverno, mas pouca durante o Verão. Por fim, o grupo a cinzento, resultante da agregação dos grupos de pequena dimensão, tem um comportamento semelhante ao grupo vermelho mas com menos incidência no Outono e mais no Inverno. A mesma análise pode ser realizada para as figuras 3.6b, 3.6c, 3.6d e 3.6e.

Para desenvolver este trabalho de agrupamento foi necessário desenvolver um programa em *python*. Foi utilizada a biblioteca *sklearn* para correr o algoritmo *Kmeans* e para obter a métrica *silhouette score*.

### 3.4.2 Espacial

Para além do agrupamento temporal foi também necessário fazer um agrupamento espacial para responder a uma das perguntas relevantes, "Existem agrupamentos espaciais de ocorrências no concelho de Lisboa?". Para responder a esta questão foi necessário recorrer a mecanismos de *clustering* com o objetivo de agrupar espacialmente as ocorrências. Este agrupamento foi feito para três conjuntos de dados. Esses conjuntos de dados são os seguintes:

- Todas as ocorrências;
- Todas as ocorrências excepto alguns "Serviços"(Fecho de Água, Abertura de Porta - Sem Socorro e Sinalizar Buraco), isto porque as ocorrências de "Serviços" representam uma elevada percentagem do conjunto de dados original;
- Todas as ocorrências de emergência. A tabela 3.12 apresenta a lista de tipologias que são consideradas de emergência pelo RSB.

Para cada um dos agrupamentos foi necessário criar conjuntos de dados. Esses conjuntos de dados apenas continham informações sobre ocorrências que pertenciam ao

Tabela 3.12: Tipologias das ocorrências consideradas de emergência.

Tipologia
Incêndios
Acidentes
Infra-estruturas e Vias de Comunicação
Serviços - Abertura de Porta - Com Socorro
Atividades - Busca/Resgate (Pessoas e Animais)

agrupamento em questão. Existia uma linha para cada ocorrência e três colunas, o identificador da ocorrência, a latitude e longitude da ocorrência.

Para o agrupamento espacial, para os vários conjuntos de dados, foi utilizado o algoritmo de *clustering DBSCAN*. A escolha deste algoritmo para fazer o agrupamento foi devido à sua capacidade de considerar ruído, pois não se pretende obrigar a que todas as ocorrências estejam num grupo, por considerar a noção de densidade e, por fim, por existir bibliotecas em *python* bem documentadas com a implementação deste algoritmo. Para utilizar este algoritmo é necessário definir dois parâmetros importantes, o *eps* e o *minPoints*. O *eps* é a distância máxima entre dois pontos para que um seja considerado próximo do outro, enquanto que o *minPoints* é o número mínimo de pontos necessários para formar um grupo. Para a escolha do melhor valor para o *eps* foi novamente utilizado o método do cotovelo. O método do cotovelo para o *DBSCAN* tenta classificar os dados e tenta encontrar a distância entre os vizinhos para encontrar a distância mínima entre eles. Isto apresenta a curva do cotovelo para encontrar a densidade dos pontos e seus valores de *eps*. A figura 3.7 apresenta o gráfico obtido para o método do cotovelo para o conjunto de dados com todas as ocorrências.

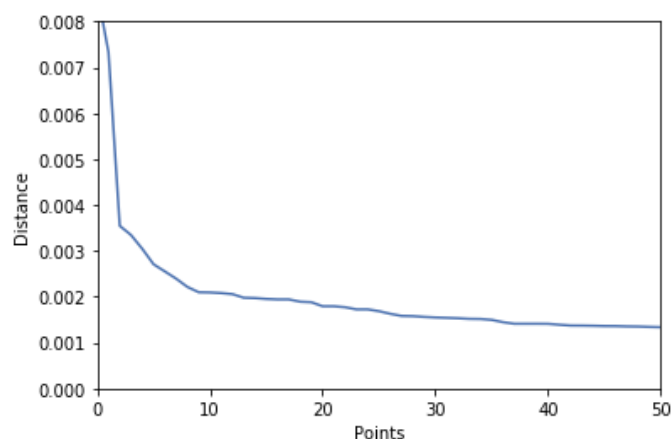


Figura 3.7: Gráfico do método do cotovelo para o conjunto de dados de todas as ocorrências do *clustering* espacial.

Como se pode verificar pela figura 3.7, o valor ideal de *eps* é 0.002, pois é onde está o cotovelo do gráfico. Depois de definido o valor de *eps* é necessário escolher o valor do *minPoints* que irá definir qual o tamanho mínimo necessário para se formar um grupo.

O conjunto de dados é extenso e os locais das ocorrências não são muito dispersos, por isso, o clássico  $minPoints = 5$  não funciona, pois assim considera quase todo o concelho de Lisboa um único grupo. Para procurar o melhor valor do  $minPoint$  foi utilizada a métrica *silhouette score*. Para encontrar o valor ideal para este parâmetro foi definido um intervalo de valores para o  $minPoints$  e um *step*. Dando o exemplo do *clustering* com todo o conjunto de dados das ocorrências de emergência, o intervalo considerado mais promissor foi o de  $[50; 100]$  com um *step* de 10, ou seja, na primeira iteração o algoritmo iria correr com o  $eps = 0.002$  e  $minPoints = 50$ , na segunda com  $eps = 0.002$  e  $minPoints = 60$  e assim sucessivamente. Para cada iteração era retirada a métrica *silhouette score* para o agrupamento resultante. Os valores dessa métrica para cada valor possível de  $minPoints$  estão apresentados na tabela 3.13.

Tabela 3.13: Valores da métrica *silhouette score* para diferentes  $minPoints$  no agrupamento com dados das ocorrências de emergência.

Valor do $minPoints$	Valor do <i>silhouette score</i>
50	-0.358
60	-0.311
70	-0.222
80	-0.133
90	-0.141
100	-0.164

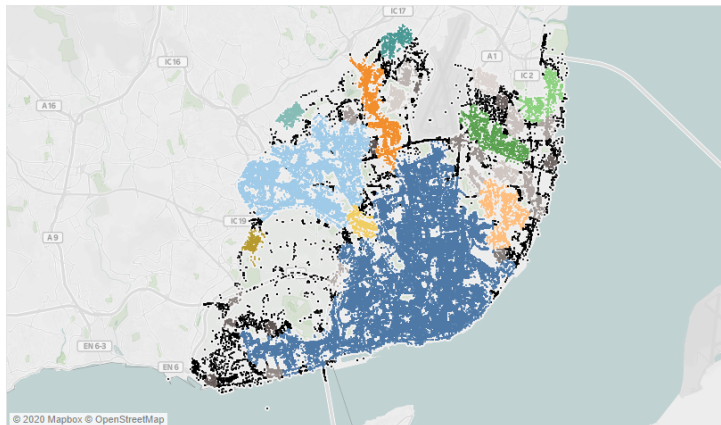
Com os valores de *silhouette score* presentes na tabela 3.13, foi escolhido o valor de 80 para o parâmetro  $minPoints$  por ser o que obtém um valor melhor da métrica. Esta metodologia foi utilizada para fazer todos os três agrupamentos espaciais. Primeiro procurar o melhor valor de  $eps$  usando o método do cotovelo e, por fim, definir um intervalo de interesse e um *step* e utilizar os valores da métrica *silhouette score* obtidos nesse intervalo para determinar qual o melhor valor para  $minPoints$ .

Tabela 3.14: Síntese do agrupamento espacial.

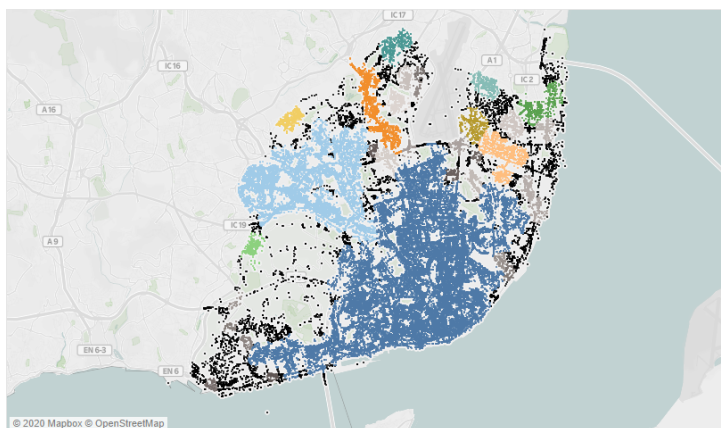
Agrupamento das ocorrências	$eps$	$minPoints$	<i>silhouette score</i>	Nº grupos
Todas	0.002	200	-0.184	33
Todas excepto alguns "Serviços"	0.002	150	-0.114	33
Todas de emergência	0.002	80	-0.133	30

A tabela 3.14 contém uma síntese do agrupamento espacial dos três conjuntos de dados considerados. Esta tabela apresenta os valores escolhidos para  $eps$  e  $minPoints$ , o valor da métrica *silhouette score* associado aos parâmetros escolhidos e o número de grupos espaciais resultantes da execução do algoritmo DBSCAN com os parâmetros escolhidos.

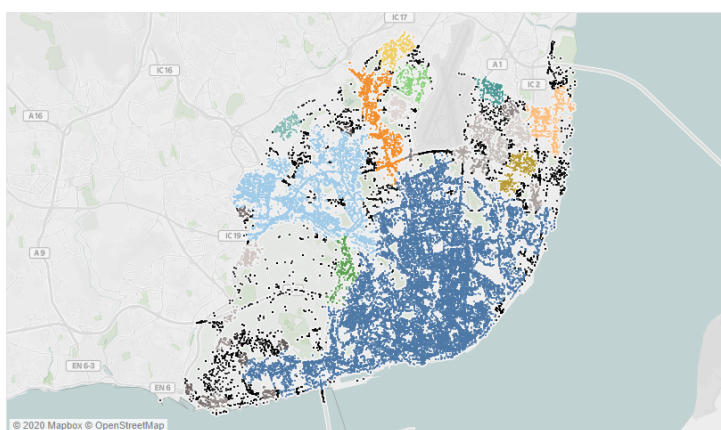
A figura 3.8 apresenta os resultados obtidos neste agrupamento espacial para os diferentes conjuntos de dados. Cada ponto no mapa representa uma ocorrência e o ponto está colorido de acordo com o grupo a que pertence. A coloração destes mapas está pensada



a Resultado com todas as ocorrências.



b Resultado com todas as ocorrências excepto alguns "Serviços".



c Resultado com as ocorrências de emergência.

Figura 3.8: Resultados do agrupamento espacial.

da seguinte forma: pontos a preto são considerados ruído, grupos coloridos são grupos que pertence ao *top* 10 de maiores dimensões e grupos numa escala de cinzentos são grupos secundários, isto é, grupos de menores dimensões que não estão presentes no *top* 10. Como podemos ver na figura 3.8 os agrupamentos resultantes não são muito diferentes entre si. Existem dois grupos grandes em todas as figuras 3.8a, 3.8b e 3.8c, o grupo a azul escuro e o grupo a azul claro.

Para o desenvolvimento deste trabalho de agrupamento foi utilizada, novamente, um programa em *python* e a biblioteca *sklearn* que contém o algoritmo do DBSCAN e a métrica *silhouette score*.

### 3.5 Dashboards Desenvolvidos

O objetivo desta dissertação é desenvolver ferramentas que permitam a exploração e análise de padrões temporais, espaciais e espaço-temporais. Durante esta dissertação foram desenvolvidos vários *dashboards* que procuram responder às questões que foram inicialmente consideradas relevantes. Estas ferramentas devem ser interativas e de fácil uso/compreensão.

Nesta seção serão apresentados os *dashboards* desenvolvidos, nomeadamente, quais as questões que eles pretendem responder, quais as visualizações e interações que contêm. Para a descrição dos *dashboards*, das suas visualizações e interações, será utilizada a linguagem gráfica IVML. Será aqui feita uma breve introdução a esta linguagem de modo a, posteriormente, ser fácil a leitura dos diagramas IVML.

O apêndice C contém todos os diagramas IVML, tanto de interações como de *tooltips*, de todos os *dashboards* que irão ser apresentados no formato de uma página inteira para maior compreensão.

#### 3.5.1 Introdução à Linguagem gráfica IVML

O IVML - *Interaction Visualization Modeling Language* é uma linguagem gráfica que pretende descrever um *dashboard*, nomeadamente, os seus componentes e interações, em diagramas de fácil leitura e compreensão. Existem dois tipos de diagramas IVML, o diagrama de interações e o diagrama de *tooltips*. O diagrama de interações apresenta a estrutura de um *dashboard*, quais as componentes existentes e quais as interações existentes nele. Este diagrama tem como objetivo descrever quais os componentes, em destaque, as visualizações, nomeadamente, qual o tipo de gráfico, quais as variáveis visuais e quais os dados que são utilizados. Também tem como objetivo apresentar quais são as interações que o *dashboard* disponibiliza aos utilizadores. Mostra qual é a ação necessária para iniciar a interação, quais os dados que são manipulados e quais as componentes/visualizações que serão afetadas. Relativamente ao diagrama de *tooltips*, este diagrama pretende apresentar quais são as *tooltips* dos diferentes componentes do *dashboard*. Os componentes

podem ser de vários tipos: imagens, texto, botões, filtros, parâmetros, legendas, *tooltips*, visualizações e *dashboard*.

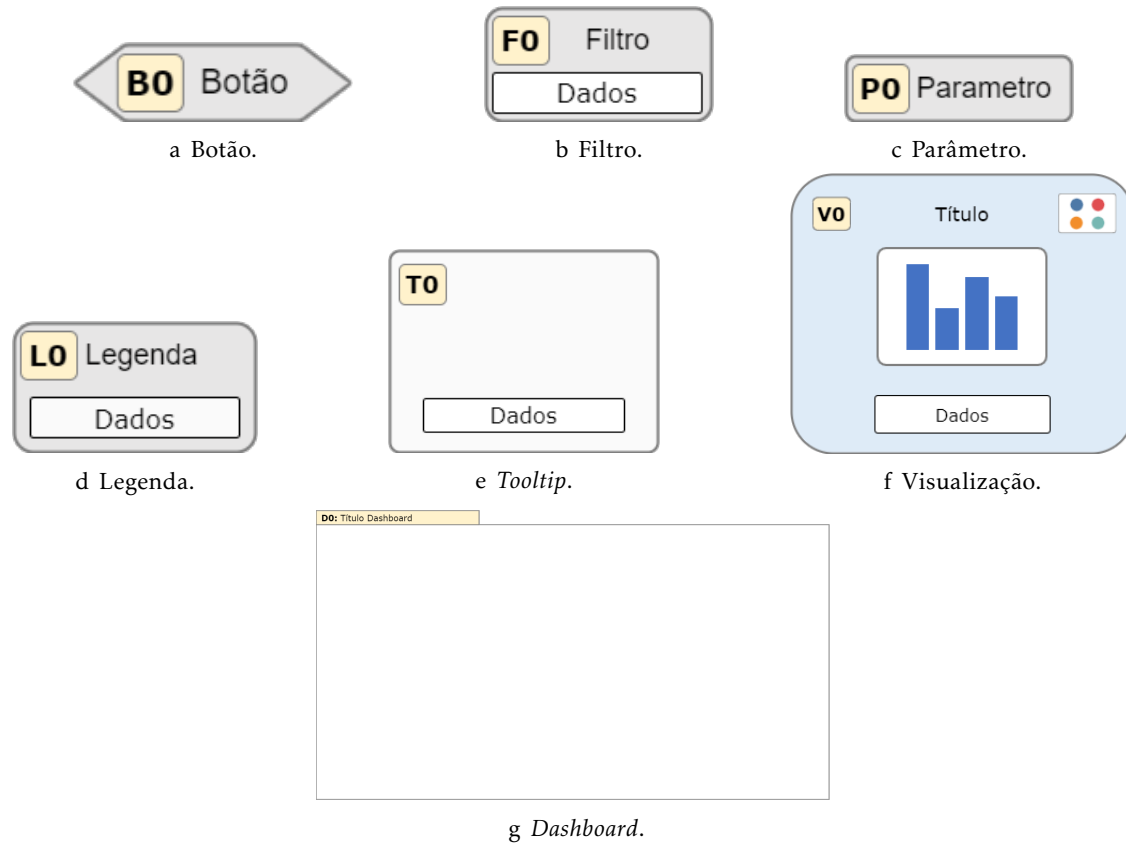


Figura 3.9: Exemplo dos componentes da linguagem IVML.

A figura 3.9 demonstra a representação gráfica de alguns dos componentes acima referidos. Como é visível na figura 3.9 todos os componentes têm um identificador na parte superior esquerda. Este identificador tem a letra inicial de acordo com o tipo de componente a que pertence. Os vários componentes podem ser identificados através da sua forma geométrica, da sua cor e da sua letra inicial do identificador. Quanto à forma geométrica, é fácil identificar componentes diferentes pela forma gráfica que apresentam, por exemplo, o componente botão presente na figura 3.9a é fácil de distinguir do componente parâmetro presente na figura 3.9c devido à sua forma, um é um hexágono e outro é um retângulo. Também é possível distinguir componentes pela cor que apresentam, por exemplo, o componente *tooltip* (na figura 3.9e) e o componente visualização (na figura 3.9f) são distinguíveis, pois um é cinzento e outro é azul. Existem componentes que não podem ser distinguidos nem pela forma nem pela cor, no caso do filtro (na figura 3.9b) e da legenda (na figura 3.9d). Nestes casos os componentes são distinguidos pela letra inicial do seu identificador, o filtro apresenta um F e a legenda apresenta um L. O componente visualmente mais complexo é o componente da visualização. Através da figura 3.9f podemos identificar que este componente pode apresentar: o identificador,





Figura 3.10: Representação de algumas interações.



Figura 3.11: Representação dos tipos de ações para iniciar as interações.

um título, quais as variáveis visuais que faz uso (canto superior direito), qual o tipo de gráfico (centro) e, por fim, quais os dados utilizados para esta visualização. Relativamente às variáveis visuais, estas podem ser: cor, tamanho e forma. Quanto ao tipo de gráficos existem várias possibilidades, por exemplo, texto, tabela, gráfico de barras, gráfico de barras *stacked*, gráfico de linhas, gráfico de área, mapas, mapa coropleto, *hexabins*, círculos, *heatmaps*, *scatter plot*, *gantt chart* e animações temporais.

Esta linguagem também permite descrever vários tipos de interações. Alguns tipos de interações mais conhecidas e utilizadas são a de filtragem e de destaque.

Estas interações são representadas na forma presente na figura 3.10, onde no centro do paralelogramo podem estar *icons* que representam o tipo de ação, por exemplo, uma filtragem (figura 3.10a) ou um destaque (figura 3.10b). As interações têm de ter um componente de partida, onde a ação é iniciada, e um ou mais componentes de chegada, componentes que serão afetados pela ação. Também é necessário ser especificado que tipo de ação é necessária para iniciar as interações, isto é, pode ser um clique num ponto da visualização ou um *hover* num ponto da visualização. A figura 3.11 demonstra a representação gráfica da ação de clique e *hover*.

A figura 3.12a apresenta um exemplo de duas interações diferentes. Uma interação de filtragem iniciada com um clique (figura 3.12a) e uma interação de destaque iniciada com um *hover* (figura 3.12b).

### 3.5.2 Comparação das Ocorrências nas Áreas de Influência dos Quartéis

O desenvolvimento de ferramentas visuais tem como objetivo criar mecanismos gráficos que ajudem os utilizadores a responder a questões relevantes. No início da dissertação foi elaborada uma lista com estas questões. Uma das perguntas consideradas relevantes foi: "Quais são as áreas de influência dos quartéis que apresentam comportamentos semelhantes/diferentes em termos de ocorrências e suas características?". No entanto, a



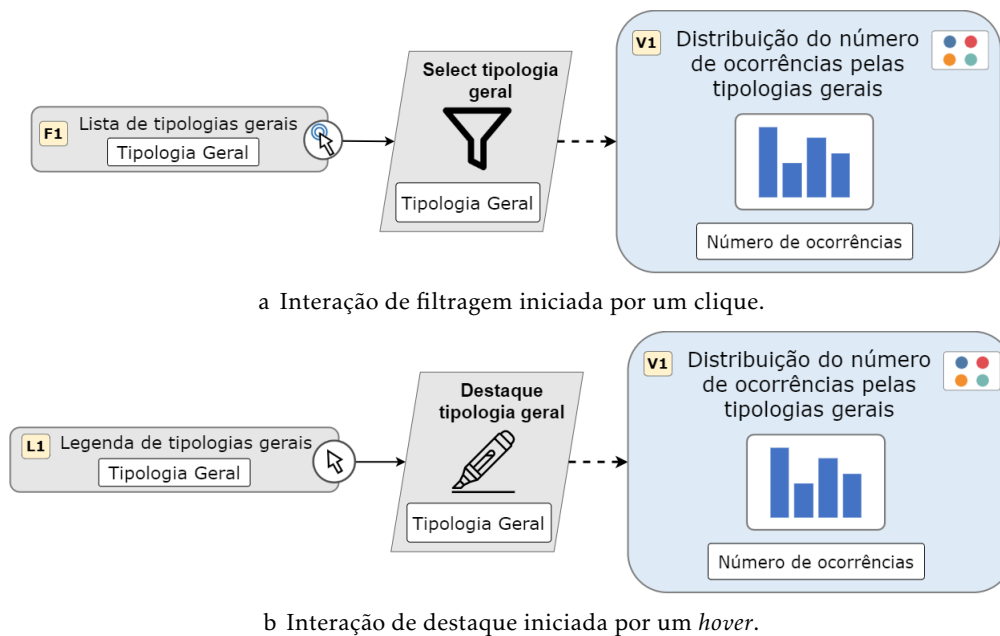


Figura 3.12: Exemplo de interações iniciadas por diferentes tipos de ações.

parte de comportamentos semelhantes/diferentes em termos de características das ocorrências é um termo vago. Por isso, foram definidas quais as características interessantes para análise. Foi definido que se devia analisar os comportamentos em termos de meios mobilizados (elementos e viaturas), das tipologias das ocorrências, da evolução temporal e da média diária de ocorrências. Para resolver esta questão foi desenvolvido o *dashboard* de comparação das ocorrências nas áreas de influência dos quartéis do RSB. Uma imagem deste *dashboard* está presente na figura 3.14 e um vídeo demonstrativo das suas funcionalidades está disponível [aqui](#).

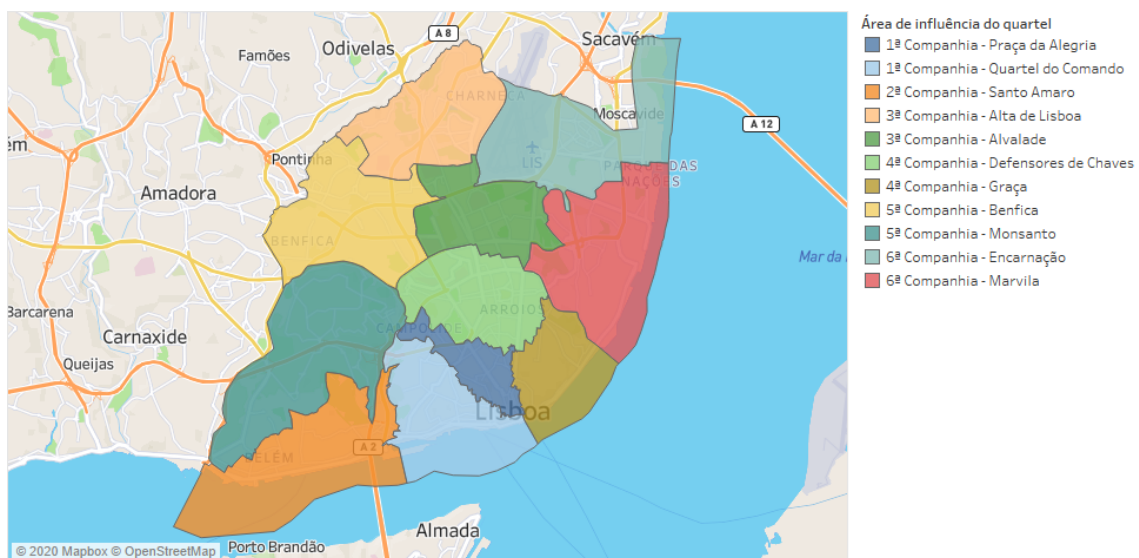


Figura 3.13: Áreas de influência dos quartéis do RSB.

A figura 3.13 representa as várias áreas de influência dos quartéis do RSB. Como é observável pela figura 3.13 existem 11 áreas de influência no concelho de Lisboa e, consequentemente, 11 quartéis do RSB.

Uma nota importante para o desenvolvimento deste *dashboard* é que não existe informação sobre qual o quartel responsável pela resposta à ocorrência. No entanto, uma ocorrência tem uma localização e é possível saber em que área de influência essa ocorrência caiu. Não vamos assumir que o quartel da área de influência onde a ocorrência caiu é o responsável pela resposta à mesma, porque pode não ser verdadeiro. Portanto, é de salientar que neste *dashboard* vai-se analisar as áreas de influência em termos de ocorrências que caíram na sua área.

Este *dashboard* foi desenhado para disponibilizar ao utilizador análises e comparações para a distribuição das ocorrências por tipologias em que o utilizador esteja interessado, média de ocorrências diárias, meios humanos e viaturas mobilizadas por ocorrências e comportamentos temporais a diferentes granularidades temporais, ao nível anual, mensal, dia da semana e horário.

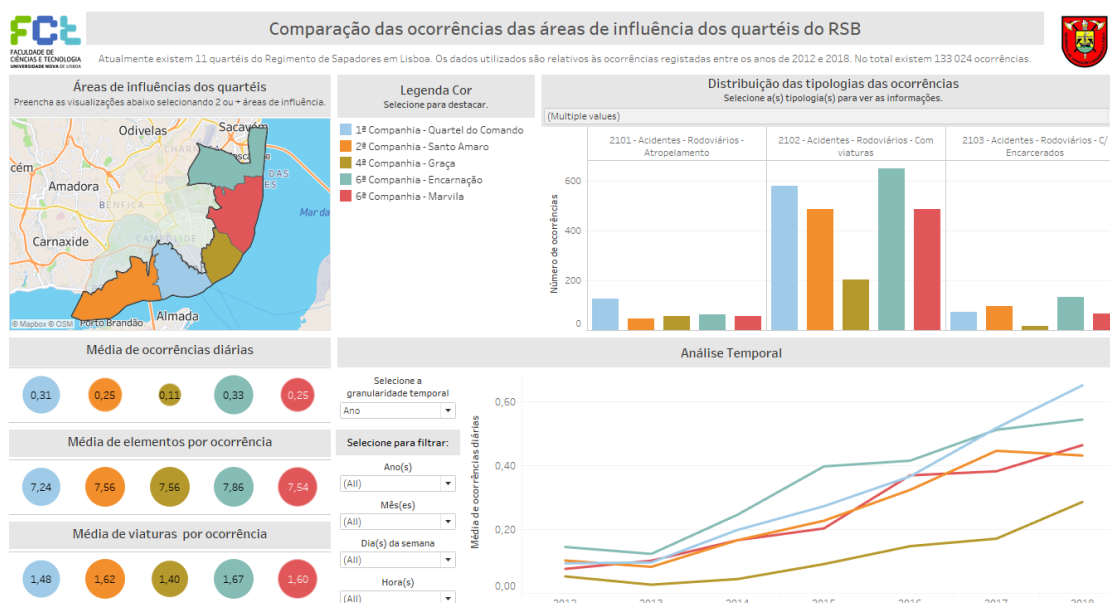


Figura 3.14: *Dashboard* de comparação das ocorrências nas áreas de influência dos quartéis do RSB.

A figura 3.14 apresenta este *dashboard* com a seleção para comparação das áreas de influência das zonas ribeirinhas. Para descrever este *dashboard* estão presentes nas figuras 3.15 e 3.17 os diagramas IVML de interações e *tooltips*, respetivamente.

Como demonstrado na figura 3.15 existem seis visualizações, uma legenda, cinco filtros e um parâmetro. A visualização V1 é um mapa coropleto que mostra quais são as áreas de influências existentes e este mapa utiliza a cor para mapear estas áreas, ou seja, cada cor corresponde a uma, e uma só, área de influência. Ao lado direito desta visualização está presente uma legenda de cor para todo o *dashboard*. Esta legenda L1

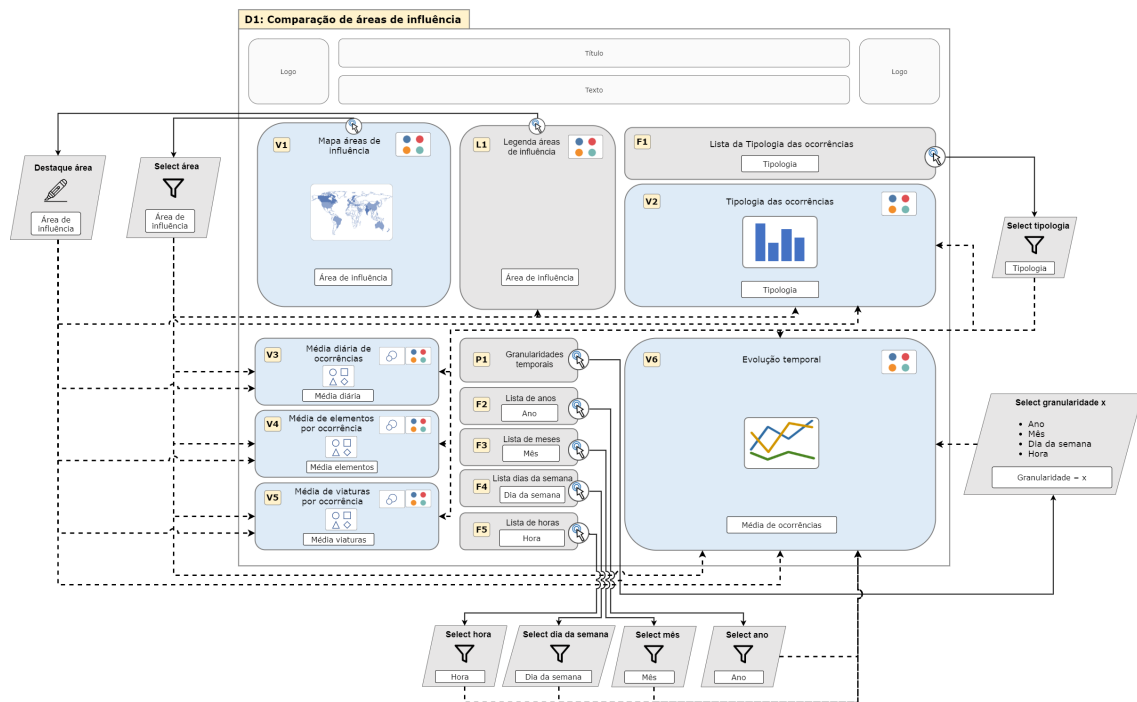


Figura 3.15: Diagrama IVML de interações do *dashboard* de comparação das ocorrências nas áreas de influência dos quartéis do RSB.

mostra o mapeamento de cor existente no *dashboard* que associa uma cor a cada uma das áreas de influência presentes no mapa da visualização V1. Também está presente um filtro, F1, que contém a listagem das várias tipologias de ocorrências que o conjunto de dados disponibiliza. Por debaixo deste filtro, está uma visualização que permite ver a distribuição das ocorrências por tipologia para cada área de influência. Esta visualização V2 é um gráfico de barras que faz uso da cor para mapear para a área de influência, ou seja, existe uma barra para cada área de influência e tipologia e essa barra terá a cor associada à área de influência. As visualizações V3, V4 e V5 são visualmente iguais, isto é, todas são gráficos de forma, nomeadamente, círculos e permitem comparar as áreas de influência em termos de média diária de ocorrências, média de elementos e viaturas por ocorrência, respetivamente. Para cada área de influência existe uma bola que terá inscrito nela um valor. Estas visualizações fazem do tamanho para a média de ocorrências diárias, média de elementos e viaturas por ocorrência, ou seja, círculos menores representam números menores e círculos maiores representam valores maiores e fazem também uso da cor, novamente, para mapear para a área de influência. O único parâmetro existente neste *dashboard* é o parâmetro P1 que diz respeito à seleção de granularidades temporais. Este parâmetro tem quatro valores possíveis: "Ano", "Mês", "Dia da semana" e "Hora". Os filtros F2, F3, F4 e F5 são, respetivamente, uma lista de anos, meses, dias da semana e horas existentes no conjunto de dados das ocorrências do RSB. Por fim, a visualização V6 apresenta a evolução temporal das ocorrências em termos de média de ocorrências ao longo do tempo. Esta visualização é um gráfico de linhas e contém uma linha para cada

uma das áreas de influência. Novamente, esta visualização utiliza a cor para mapear para as áreas de influência.

Relativamente às interações disponíveis neste *dashboard*, existem seis interações de filtragem, uma de destaque e uma de escolha de parâmetros. A interação mais importante neste *dashboard* é a seleção das áreas de influência que estamos interessados em comparar. Para esta seleção podem ser escolhidas duas ou mais áreas de influência de interesse. Esta escolha é feita através da seleção (via clique) das áreas de influência presentes na visualização V1, o mapa coropleto. Esta seleção vai afetar todas as outras visualizações do *dashboard*, incluindo a legenda de cores. Os componentes afetados passarão apenas a apresentar dados relativos às áreas de influência selecionadas. Existe um filtro para filtrar quais as tipologias a considerar no *dashboard*. Este filtro é acionado quando se seleciona, ou se retira da seleção, tipologias presentes na lista de tipologias do filtro F1. Este filtro afeta as visualizações V2, V3, V4, V5 e V6. A figura 3.16 apresenta o filtro F1 onde é possível selecionar e remover quaisquer tipologias existentes no conjunto de dados das ocorrências. Deste modo, é possível analisar e comparar as áreas de influências apenas nas tipologias que têm interesse de análise. Existe uma ação de escolha de parâmetro e esta ação é iniciada quando se seleciona a granularidade temporal de interesse. Esta ação apenas afeta a visualização V6, evolução temporal. Este parâmetro permite que a análise da evolução temporal seja feita em várias granularidades temporais. No entanto, em algumas delas, como é o caso dos meses, dias da semana e horas, existe uma dobra temporal. Como estamos a funcionar com vários anos de dados, quando se muda para a granularidade temporal "Mês" estamos a apresentar dados de vários Janeiros, Fevereiro e etc. Assim introduzimos uma dobra temporal na análise. Para contornar esta questão, estão disponíveis os filtros F2, F3, F4 e F5. Ao se adicionar, ou retirar, valores destas listas estamos a afetar a visualização V6 da evolução temporal e, se for o desejado, é possível analisar os dados sem a existência da dobra temporal. Por fim, está disponível uma interação de destaque neste *dashboard*. Esta ação pode ser iniciada através de um clique na área de influência presente na legenda de cores L1. Ao se clicar na área que se pretende destacar, todas as visualizações V2, V3, V4, V5 e V6 serão afetadas, destacando apenas os valores referentes à área selecionada.

Relativamente às *tooltips*, como podemos ver pela figura 3.17 todas as *tooltips* são textuais. A *tooltip* da visualização V1, mapa coropleto com as áreas de influência, é apresentada sempre que se faz *hover* do rato numa das áreas de influência e são apresentadas informações sobre o número de ocorrências que caíram nessa área e a percentagem do total de ocorrências, isto é, a percentagem de ocorrências que a área de influência tem tendo em conta todas as ocorrências do conjunto de dados. As *tooltips* T2, T3 e T4 são relativas às visualizações V3, V4 e V5, respetivamente. Estas *tooltips* apresentam apenas os valores que já são mostrados na visualização, a média diária de ocorrências, a média de elementos e viaturas enviadas por ocorrências da área de influência e se o filtro da tipologia estiver ativo estes valores estão também filtrados. As informações aparecem no

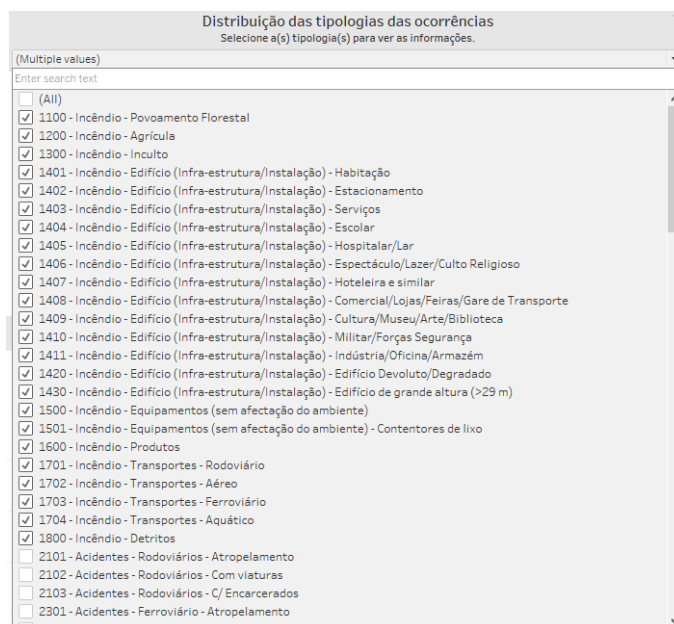


Figura 3.16: Exemplo do filtro F1 que contém uma lista com todas as tipologias das ocorrências do RSB e é possível selecionar ou remover tipologias para filtrar os dados.

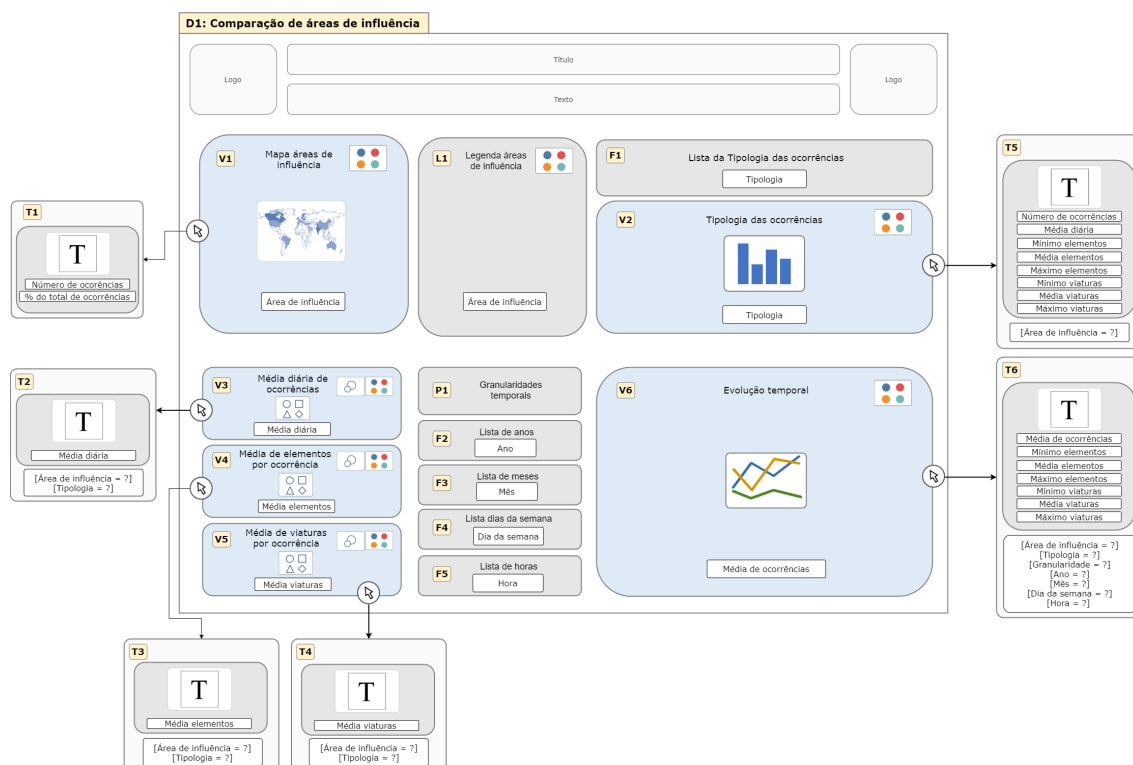


Figura 3.17: Diagrama IVML de *tooltips* do *dashboard* de comparação das ocorrências nas áreas de influência dos quartéis do RSB.

ecrã quando se faz *hover* com o rato num dos círculos das visualizações. A *tooltip* T5 pertence à visualização V2, distribuição das ocorrências por tipologia, e aparece no *dashboard* sempre que se faz *hover* do rato numa das barras do gráfico relativas à tipologia na área de influência. Esta *tooltip* apresenta as seguintes informações sobre as ocorrências daquela tipologia naquela área de influência: número de ocorrências, média diária de ocorrências, valores mínimo, médio e máximo de elementos e viaturas mobilizadas em ocorrências da tipologia na área de influência. Por fim, a *tooltip* da visualização V6, da evolução temporal, é ativada quando se faz um *hover* do rato num ponto do gráfico de linhas. São apresentadas as seguintes informações na *tooltip* T6: média de ocorrências, valor mínimo, médio e máximo de elementos e viaturas mobilizadas para o local da ocorrência. Normalmente, toda a informação apresentada está pendente dos filtros que afetam esta visualização, nomeadamente, a área de influência, as tipologias, os anos, meses, dias da semana e horas. A título de exemplo, a figura 3.18 apresenta uma imagem da *tooltip* T5.

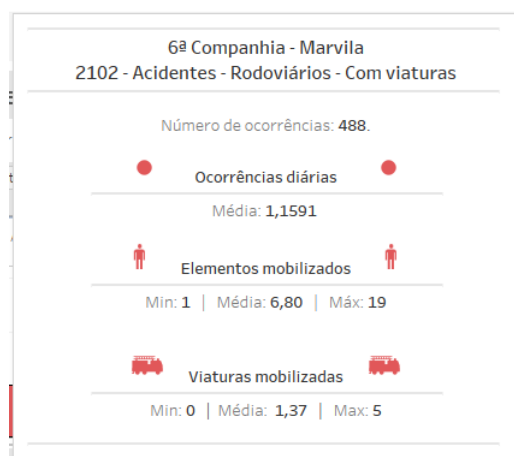


Figura 3.18: Exemplo de uma *tooltip*, *tooltip* T5 do *dashboard* de comparação das ocorrências nas áreas de influência dos quartéis do RSB.

Para concluir e resumir, com este *dashboard* é possível:

1. Saber quais são as áreas de influência disponíveis e que área geográfica ocupam. Isto permite responder a questões como "Quantos quartéis existem? Qual a área de influência dos mesmos?";
2. Analisar a distribuição das ocorrências por tipologia de cada área de influência. Isto permite observar se existem áreas de influência com mais incidência numa certa tipologia que outras áreas. Assim é possível responder a perguntas como "Qual a distribuição do número de ocorrências pelas diversas tipologias nas áreas de influência?" ou "Existe alguma área de influência com maior incidência de um certo tipo de tipologia que outra área?";
3. Comparar as áreas de influência em termos de meios mobilizados para o local,

elementos e viaturas. Isto permite saber se existem áreas que dispensam muitos/-poucos meios em relação a outras áreas. Assim é possível responder às perguntas "Em média, qual das áreas de influência que estou a analisar envia mais elementos/-viaturas por ocorrência?";

4. Comparar a incidência de ocorrências das áreas de influência, isto é, se uma área de influência tem, em média, mais ocorrências por dia do que outra área. Isto permite responder a questões como "Qual a área de influência em análise contém, em média, mais ocorrências diárias?";
5. Comparar o comportamento temporal das ocorrências das áreas de influência em diferentes granularidades temporais. Isto permite saber se existem áreas de influência que partilham o mesmo perfil temporal ou não. Isto permite responder a questões como "Quais das áreas de influência em análise têm um comportamento temporal semelhante? Isso acontece para todas as granularidades temporais?".

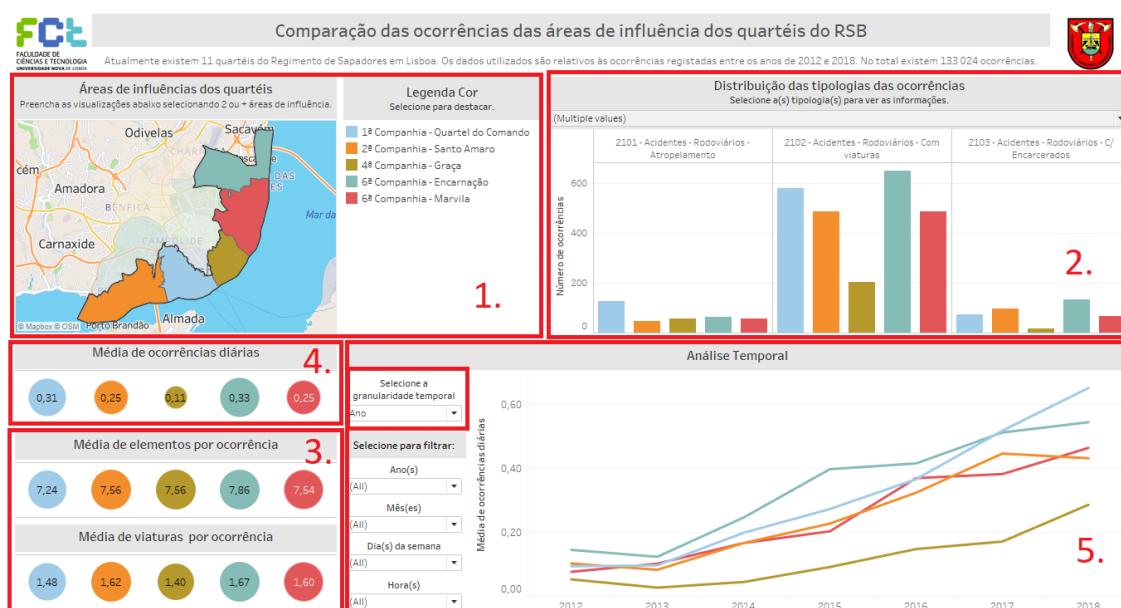


Figura 3.19: Imagem do *dashboard* de comparação das ocorrências nas áreas de influência dos quartéis do RSB com indicação das visualizações que respondem às questões.

A figura 3.19 apresenta quais as visualizações deste *dashboard* que permitem saber os pontos enumerados acima. Cada número presente na figura corresponde ao número do enumerado. Desta forma é possível fazer um mapeamento entre as visualizações e algumas das questões (enumeradas acima) que estas são capazes de responder.



### 3.5.3 Comparação das Ocorrências nas Freguesias de Lisboa

A comparação das várias freguesias do concelho de Lisboa, em termos de ocorrências, foi considerada uma análise também importante. Na lista de questões relevantes apresentada na seção 3.1, está presente a questão: "Quais são as freguesias que apresentam comportamentos semelhantes/diferentes em termos de ocorrências e suas características?". O *dashboard* da comparação das ocorrências nas freguesias de Lisboa pretende responder a esta questão. Tal como apresentado na subseção acima, a questão é vaga e por isso foi considerado que as freguesias têm de ser analisadas em termos de meios mobilizados para o local da ocorrência (elementos e viaturas), das tipologias, do comportamento temporal e da média diária de ocorrências. Essencialmente este *dashboard* permite abordar os mesmos tópicos que o *dashboard* anteriormente apresentado e como tal as suas estruturas são semelhantes. Uma imagem deste *dashboard* está presente na figura 3.20 e um vídeo demonstrativo das suas funcionalidades está disponível também [aqui](#).

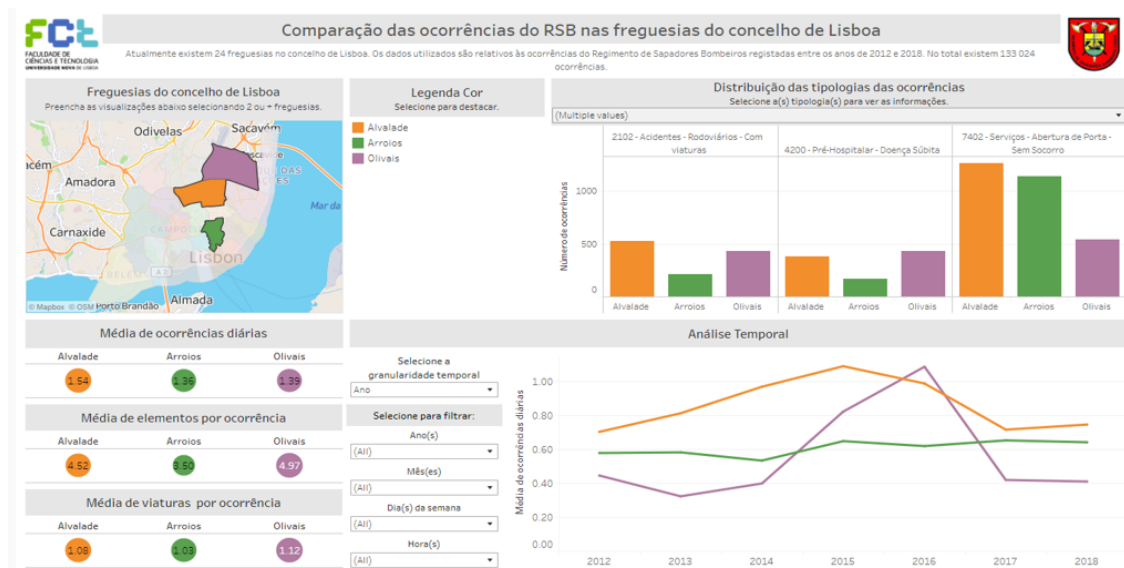


Figura 3.20: *Dashboard* de comparação das ocorrências nas freguesias do concelho de Lisboa.

Como é visível pela figura 3.20 este *dashboard* é similar ao *dashboard* apresentado na subseção anterior, o da comparação das ocorrências nas áreas de influência dos quartéis do RSB. Como a análise é semelhante e apenas se muda a organização administrativa das áreas, isto é, deixam de ser áreas de influência e passam a ser freguesias, as visualizações e interações permanecem as mesmas. A única diferença entre estes dois *dashboards* são os dados utilizados. Para referir que ao contrário do *dashboard* anterior, neste caso, sabemos exatamente a que freguesia a ocorrência pertence.

As figuras 3.21 e 3.22 apresentam os diagramas IVML de interações e *tooltips*, respetivamente.

Seguindo a figura 3.21, as componentes existentes neste *dashboard* são as mesmas que existiam no *dashboard* anteriormente apresentado, excepto a visualização V1 e a legenda



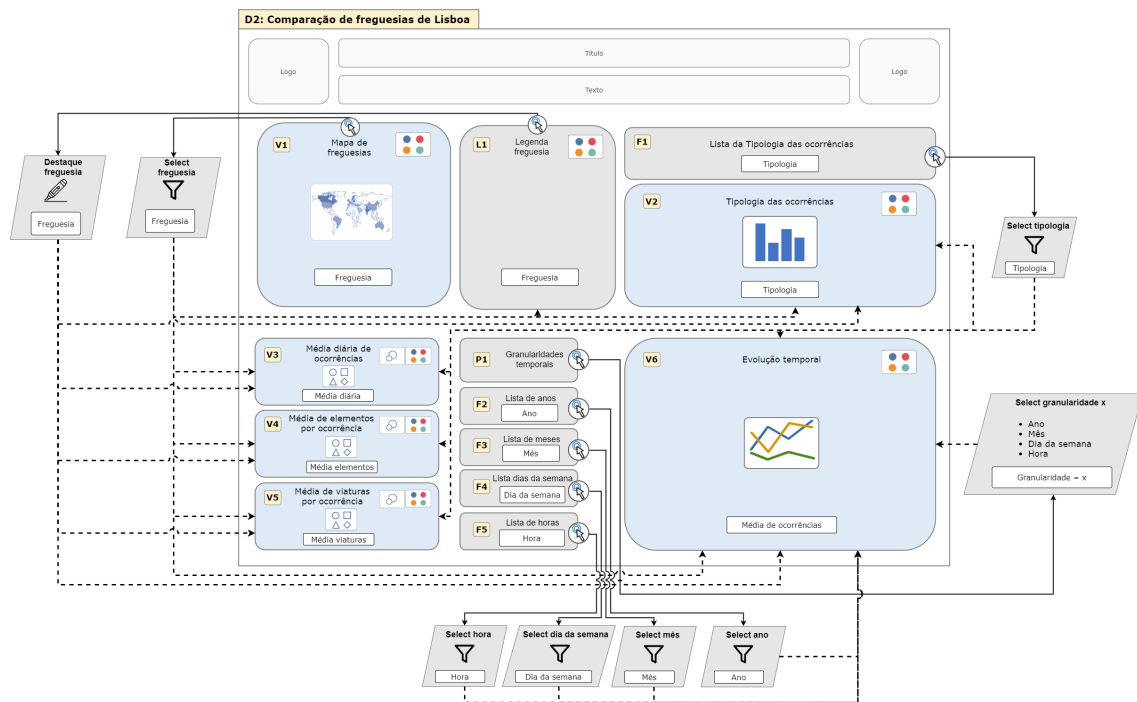


Figura 3.21: Diagrama IVML de interações do *dashboard* de comparação das ocorrências nas freguesias do concelho de Lisboa.

L1 que utilizam dados diferentes. Existe um mapa coropleto (V1) que demonstra as áreas ocupadas pelas freguesias de Lisboa e cada uma das freguesias tem uma cor associada que irá identificar essa freguesia em todo o *dashboard*. Existe uma legenda de cores que faz essa associação de uma freguesia a uma só cor identificativa (L1). Há um filtro que é igual ao presente na figura 3.16. Este filtro (F1) apresenta todas as tipologias das ocorrências do RSB. Por debaixo do filtro F1 existe um gráfico de barras (V2) que apresenta a distribuição do número de ocorrências ao longo das tipologias que aconteceram nas freguesias, sendo que estas são identificadas pela cor da barra. As visualizações V3, V4 e V5 apresentam dados como a média de ocorrências diárias, média de elementos e viaturas enviados por ocorrência, respetivamente. Os dados são apresentados em formato de círculos cujo tamanho está mapeado para o valor da média a que se refere e a cor para a freguesia. O parâmetro da granularidade temporal (P1) dispõe do mesmo conjunto de granularidades temporais que o *dashboard* anterior, anos, meses, dias da semana e horas. Os filtros F2, F3, F4 e F5 também dispõem do mesmo conjunto de dados que o *dashboard* anterior. Estes filtros servem para o utilizador desfazer a dobra temporal existente, se assim o pretender. Por fim, o *dashboard* apresenta um gráfico de linhas para análise temporal (V6) da média de ocorrências diárias.

Continuando a seguir a figura 3.21, as interações existentes neste *dashboard* são novamente semelhantes às existentes no *dashboard* anterior. A interação mais importante volta a ser a seleção, neste caso, do conjunto de freguesias de interesse através da visualização V1. Esta ação irá afetar todo o *dashboard*. Existe também a ação de destaque das freguesias.

Esta ação é iniciada através da seleção das freguesias a destacar no componente L1 e afeta todas as outras visualizações. As ações de filtragem são as mesmas que anteriormente apresentadas. Neste caso não existem diferenças entre as anteriores descritas, pois os dados que serão utilizados para proceder à filtragem são iguais. Por fim, a ação de seleção de parâmetro volta também a ser igual à anteriormente descrita. O utilizador irá selecionar a granularidade temporal desejada do conjunto existente e esta escolha afetará apenas a visualização V6 da análise temporal.

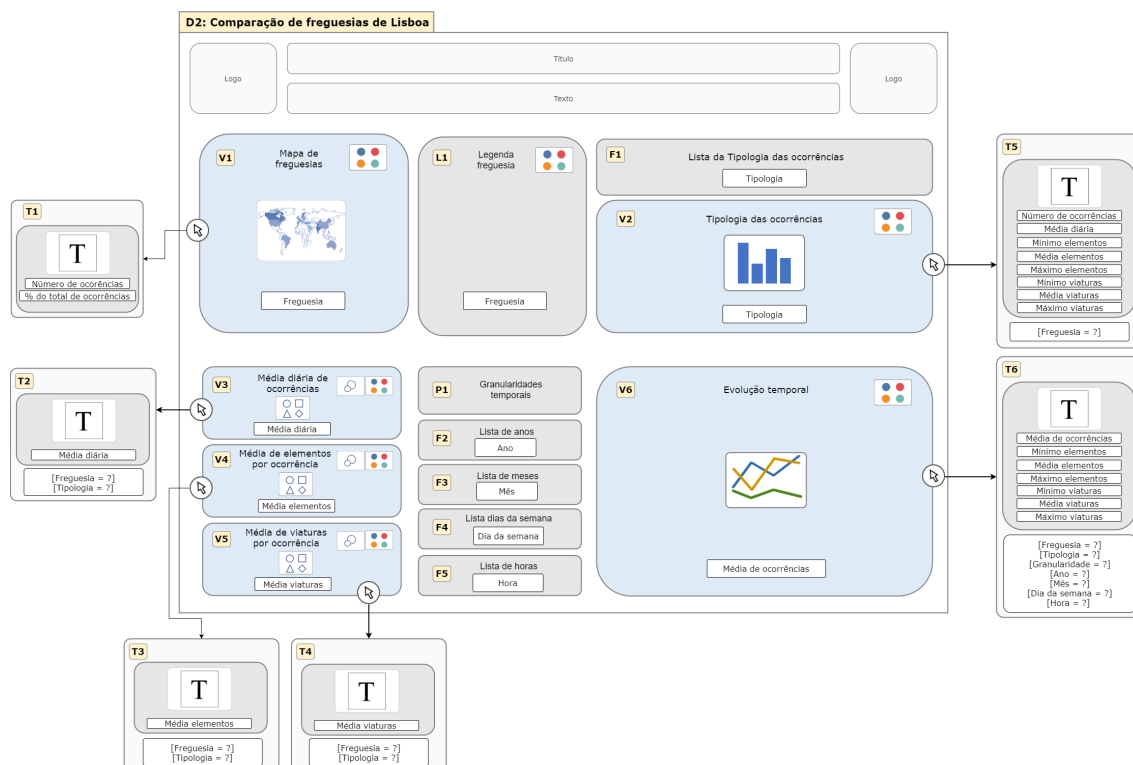


Figura 3.22: Diagrama IVML de tooltips do dashboard de comparação das ocorrências nas freguesias do concelho de Lisboa.

Relativamente às *tooltips*, o diagrama presente na figura 3.22 mostra que as *tooltips* são iguais às do dashboard de comparação das ocorrências nas freguesias de Lisboa. As *tooltips* deste dashboard apenas são diferentes no aspecto que serão filtradas de acordo com as freguesias escolhidas e não as áreas de influência.

Para concluir e resumir, com este dashboard é possível:

1. Saber quais são as freguesias do concelho de Lisboa e qual a área geográfica que ocupam, permitindo assim responder a questões como "Quais são as freguesias do concelho de Lisboa? Onde elas se localizam?";
2. Analisar a distribuição das ocorrências por tipologia de cada freguesia. Isto possibilita a resposta a perguntas como "Qual a distribuição do número de ocorrências pelas diversas tipologias das freguesias em análise?" e "Existe alguma das freguesias com mais incidência de um certo tipo de ocorrências?";

3. Comparar as freguesias em termos de meios mobilizados para o local, elementos e viaturas. Permite obter resposta à questão "Qual das freguesias em análise tem maior número de elementos e viaturas mobilizadas por ocorrência?";
4. Comparar as freguesias em termos de média diária de ocorrências. Permite responder à questão "Qual das freguesias em análise tem mais ocorrências, em média, por dia?";
5. Comparar o comportamento temporal das ocorrências nas várias freguesias em diferentes granularidades temporais. Assim é possível "Quais das freguesias apresentam um comportamento temporal semelhante?".

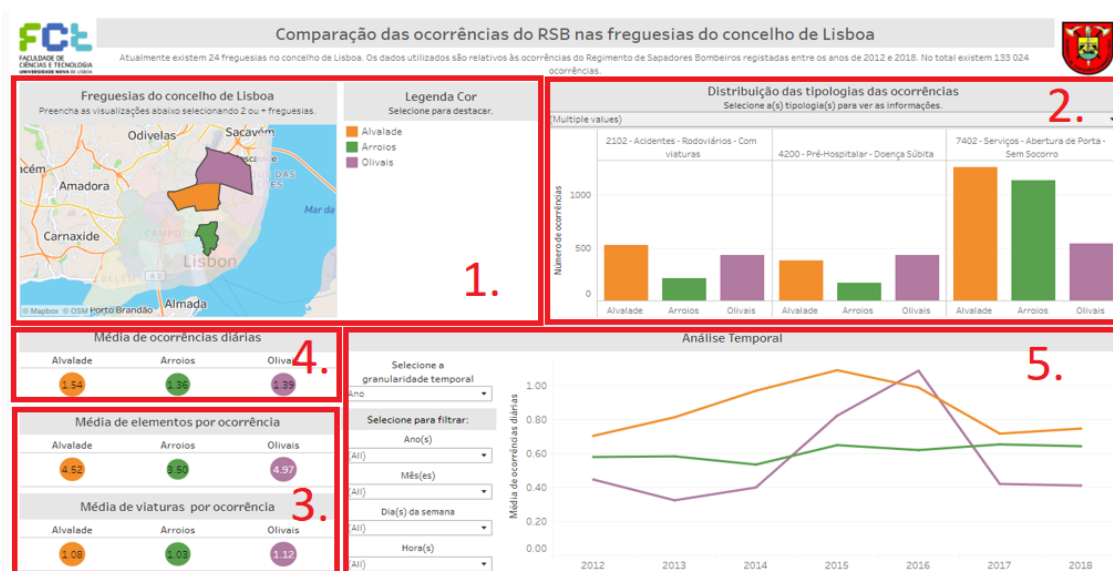


Figura 3.23: Imagem do *dashboard* de comparação das ocorrências nas freguesias de Lisboa com indicação das visualizações que respondem às questões.

A figura 3.23 apresenta quais as visualizações deste *dashboard* que permitem saber os pontos enumerados acima. Cada número presente na figura corresponde ao número do enumerado. Desta forma é possível fazer um mapeamento entre as visualizações e algumas das questões (enumeradas acima) que estas são capazes de responder.

### 3.5.4 Evolução Espacial das Tipologias das Ocorrências ao Longo do Tempo

Os *dashboards* anteriormente apresentados não respondem a todas as questões relevantes que se levantaram inicialmente. As questões "Existem *hotspots* de ocorrências de certas tipologias no concelho de Lisboa? Esses *hotspots* movem-se ao longo do tempo?" ainda não foram respondidas, por isso, foi desenvolvido um novo *dashboard* que permite obter resposta a estas perguntas. O conceito "ao longo do tempo" é abrangente e foi definido que este *dashboard* iria permitir analisar os *hotspots* ao longo dos anos, dos trimestres e dos dias da semana. Assim é possível ver a evolução espacial das ocorrências ao longo de várias

granularidades temporais, anos, trimestres e dias da semana. Deste modo, foi necessário desenvolver um *dashboard* para cada granularidade temporal utilizada. Uma imagem de um desses *dashboards*, nomeadamente da análise anual, está presente na figura 3.24 e um vídeo demonstrativo das funcionalidades de todas as análises, independentemente da granularidade temporal, está disponível [aqui](#).

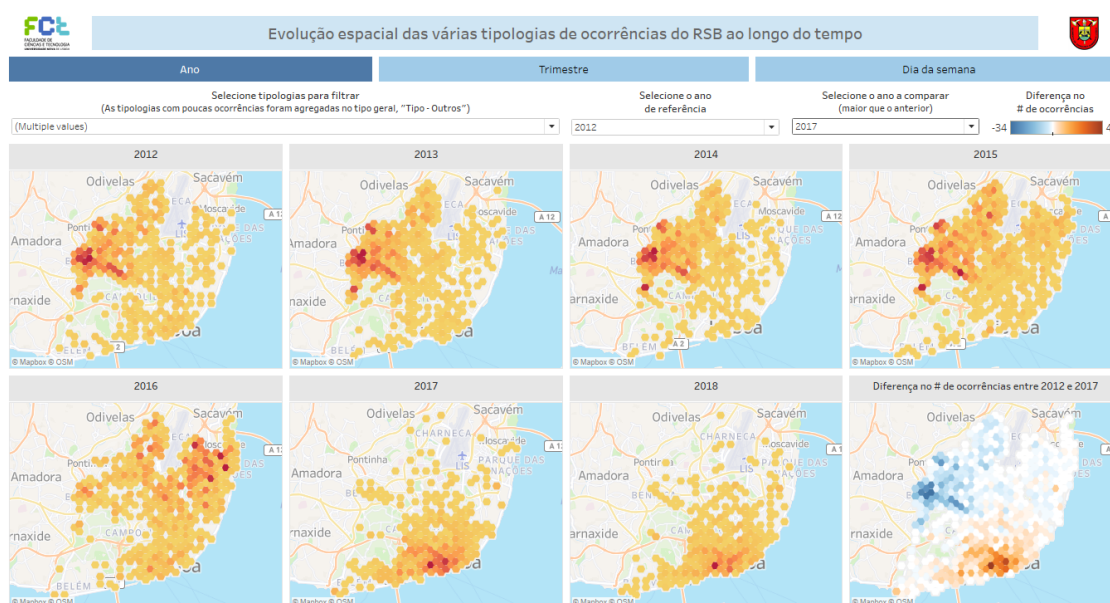


Figura 3.24: *Dashboard* da evolução espacial das tipologias das ocorrências ao longo dos anos.

Para o desenvolvimento destes *dashboards*, cujo objetivo é detetar *hotspots* e analisar se esses se movem ao longo do tempo, foi necessário definir como seriam apresentadas as visualizações para as diferentes granularidades temporais. Os dados contêm 7 anos de ocorrências, portanto, para avaliar a existência de *hotspots* ao longo dos anos tomou-se a decisão de ter 7 visualizações disponíveis ao mesmo tempo onde cada uma estava responsável por apresentar os dados de um determinado ano. Desta forma o utilizador era capaz de visualizar o que se passou de ano para ano. O mesmo pensamento foi utilizado para as restantes granularidades temporais. Relativamente aos trimestres, inicialmente a ideia era apresentar os dados ao nível do mês, mas sendo que existem 12 meses e colocar 12 visualizações no mesmo espaço não é uma opção razoável essa ideia foi sacrificada e foi utilizada a granularidade temporal de trimestres. Existem quatro trimestres, por isso só seriam necessárias 4 visualizações para apresentar a evolução espacial ao longo dos trimestres. Por último, a granularidade temporal dias da semana segue a mesma linha de pensamento. Para analisar a evolução espacial ao longo dos dias da semana seriam também necessárias 7 visualizações. Outra ideia que surgiu de nestes *dashboards* foi a de existir uma visualização final que ajudaria o utilizador a fazer a comparação entre dois anos. Por fim, toda esta análise deverá poder ser realizada para cada tipologia de ocorrências de interesse.

As figuras 3.25 e 3.26 apresentam os diagramas IVML de interações e *tooltips* do *dashboard* que permite visualizar a existência e movimento de *hotspots* ao longo dos anos, respectivamente.

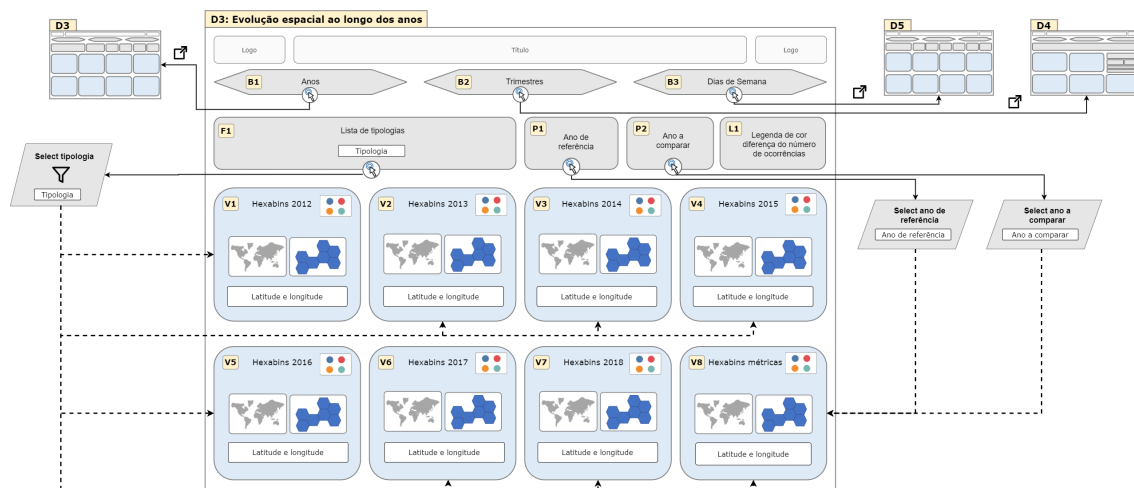


Figura 3.25: Diagrama IVML de interações do *dashboard* da evolução espacial das tipologias das ocorrências ao longo dos anos.

Como é possível observar pela figura 3.25, o *dashboard* é composto por três botões, um filtro, dois parâmetros, uma legenda e oito visualizações. Os botões presentes neste *dashboard* servem para navegação entre os outros *dashboards* que apresentam a análise em diferentes granularidades temporais. O filtro F1 apresenta a listagem de algumas tipologias. Este filtro não apresenta todas as tipologias existentes no conjunto de dados das ocorrências do RSB, porque irá servir para filtrar visualizações e algumas tipologias têm poucos registros e isto irá influenciar as visualizações e consequentemente as análises. As tipologias têm um tipo geral e subtipos associados. Portanto, as tipologias do mesmo tipo geral que apresentem poucos registros de ocorrências foram agrupadas e foi formada uma nova tipologia "Designação do Tipo Geral - Outros". As tipologias com muitos registros e o conjunto de tipologias agrupadas são listadas neste filtro F1. Os parâmetros P1 e P2 são listas que apresentam todos os anos disponíveis no conjunto de dados e representam, respectivamente, o ano de referência e o ano a comparar. Estes parâmetros estão presentes para facilitar o utilizar a comparar dados entre dois anos à escolha. A legenda L1 é uma legenda de cor relativa à visualização V8 e mostra a escala de cor e valores referentes à diferença no número de ocorrências entre dois anos. Relativamente às visualizações, conceptualmente as visualizações V1 até à V7 são iguais, isto é, todas são mapas de *hexabins* que utilizam a cor para colorir os hexágonos de acordo com o número de ocorrências presentes neles. Para cada ano de dados existe um mapa com esta descrição, ou seja, existem sete anos de dados logo existem sete mapas de *hexabins*. Cada mapa de *hexabins* apresenta apenas dados de um determinado ano. A visualização V1 apresenta dados das ocorrências de 2012 (primeiro ano que tem registros), a V2 apresenta dados de 2013 e etc. A escolha de mapas de *hexabins* passou por os hexágonos pavimentarem bem uma área geográfica. Por

exemplo, se fossem círculos iriam existir espaços entre eles, pois não existe forma de os arranjar sem deixar espaços vazios entre eles. Os hexágonos têm também vantagens sobre os quadrados, porque têm a mesma distância até aos centros dos hexágonos adjacentes e nos quadrados isso não acontece. Por fim, a visualização V8 é também um mapa de *hexabins* que utiliza a cor, mas tem uma função diferente das restantes visualizações. Esta visualização pretende demonstrar ao utilizador a diferença entre dois determinados anos em termos do número de ocorrências. Esta diferença no número de ocorrências é mapeada para a cor, portanto, os hexágonos serão coloridos de acordo com este atributo. Com esta visualização é possível observar quais são as áreas onde existiu um aumento ou uma diminuição no número de ocorrências. A visualização V8 é controlada pelos parâmetros P1 e P2 e tem o intuito de servir como resumo das restantes visualizações.

Relativamente às interações, seguindo a figura 3.25, existem três ações de navegação, uma de filtragem e duas de escolha de parâmetros. As ações de navegação são referentes aos botões presentes no *dashboard*. Ao se clicar no botão B1 volta-se a este *dashboard*, ao se clicar no B2 vai-se para o *dashboard* que apresenta a análise ao nível dos trimestres e, por fim, ao se clicar no B3 vai-se para o *dashboard* que apresenta a análise ao nível dos dias da semana. Quanto à filtragem, esta inicia-se quando se seleciona ou retira valores da lista de tipologias presente no filtro F1. Esta ação afeta todas as visualizações presentes do *dashboard*, do V1 até ao V6. As visualizações afetadas irão apenas utilizar os dados das ocorrências cujas tipologias estão selecionadas em F1. Quanto às ações de escolha dos parâmetros, estas iniciam-se com a seleção de um dos valores possíveis presentes nos parâmetros P1 e P2. A escolha de um dos valores nestes parâmetros irá afetar apenas a visualização V8. A visualização V8 irá então apresentar, em forma de um mapa de *hexabins*, a diferença no número de ocorrências entre os anos selecionados nesses parâmetros permitindo, assim, visualizar as áreas onde houve uma diminuição/aumento no número de ocorrências.

A figura 3.26 apresenta as *tooltips* presentes neste *dashboard*. Como se pode analisar, apenas existem dois tipos de *tooltips*. A *tooltip* T1 é partilhada por todas as visualizações excepto a V8. Esta *tooltip* tem como objetivo apresentar a evolução temporal das ocorrências. A T1 aparece no ecrã quando se faz *hover* do rato num hexágono das visualizações V1 até V7 (inclusive) e apresenta os seguintes dados relativamente às ocorrências que aconteceram naquela área ocupada pelo hexágono: a latitude e longitude do centro do hexágono, número de ocorrências e a percentagem do total do número de ocorrências desse hexágono, tudo isto em forma textual, depois são apresentados três gráficos de linhas que têm no eixo vertical o número de ocorrências e no eixo horizontal a granularidade temporal mês, dia de semana e hora, sendo, assim, possível visualizar o comportamento temporal de um hexágono. As informações presente na *tooltip* T1 estão dependentes das tipologias selecionadas no filtro F1. Um exemplo desta *tooltip* está presente na figura 3.27. Por fim, a *tooltip* T2 aparece no ecrã quando se faz *hover* do rato num hexágono da visualização V8. Esta apresenta apenas o valor da diferença no número de ocorrências do hexágono. Os dados apresentados nesta *tooltip* está também dependente das tipologias que estão



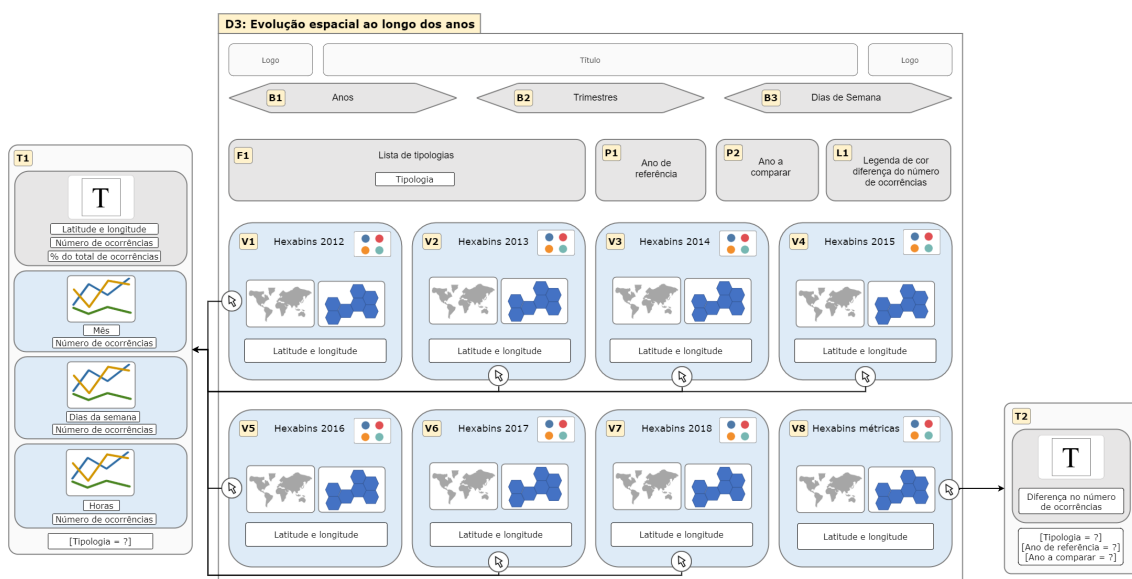


Figura 3.26: Diagrama IVML de tooltips do dashboard da evolução espacial das tipologias das ocorrências ao longo dos anos.

acionadas no filtro F1 e depende dos anos que foram selecionados nos parâmetros P1 e P2.

Como referido anteriormente, para cada granularidade temporal existe um *dashboard* diferente. A figura 3.28 apresenta uma imagem deste *dashboard*, mas para a granularidade temporal de trimestres. As figuras 3.29 e 3.30 apresentam os diagramas IVML de interações e tooltips do *dashboard* que permite visualizar a existência e movimento de *hotspots* ao longo dos trimestres, respetivamente.

Como ilustrado na figura 3.29, este *dashboard* é similar em alguns aspectos ao *dashboard* apresentado anteriormente. Este *dashboard* contém os mesmos botões B1, B2 e B3 que servem para navegação entre *dashboards*. O filtro F1 apresenta a listagem de algumas tipologias das ocorrências. Este filtro é igual ao filtro F1 descrito anteriormente, apenas apresenta algumas tipologias, as que têm mais registos, e as restantes são agrupadas no seu tipo geral. Também existem dois parâmetros que servem para controlar a visualização V5. Estes parâmetros são o P1 e P2 que apresentam uma lista de trimestres presentes no conjunto de dados das ocorrências do RSB. O P1 é o trimestre de referência e o P2 é o trimestre que se pretende comparar com o trimestre selecionado em P1. A legenda L1 é a legenda de cor da visualização V5 que mostra o gradiente de cor utilizado para mapear a diferença no número de ocorrências entre dois trimestres. Relativamente às visualizações, as visualizações de V1 até V4 (inclusive) são mapas de *hexabins* que utilizam a cor para colorir os hexágonos de acordo com o número de ocorrências presentes neles. Existe um desses mapas para cada trimestre presente nos dados. Portanto, sendo que existem quatro trimestres existem também quatro mapas de *hexabins* e cada um deles apenas contém dados de ocorrências do trimestre que representa. A última visualização é a V5 que também é um mapa de *hexabins* que faz uso da variável visual cor para colorir os hexágonos de

## CAPÍTULO 3. TRABALHO REALIZADO

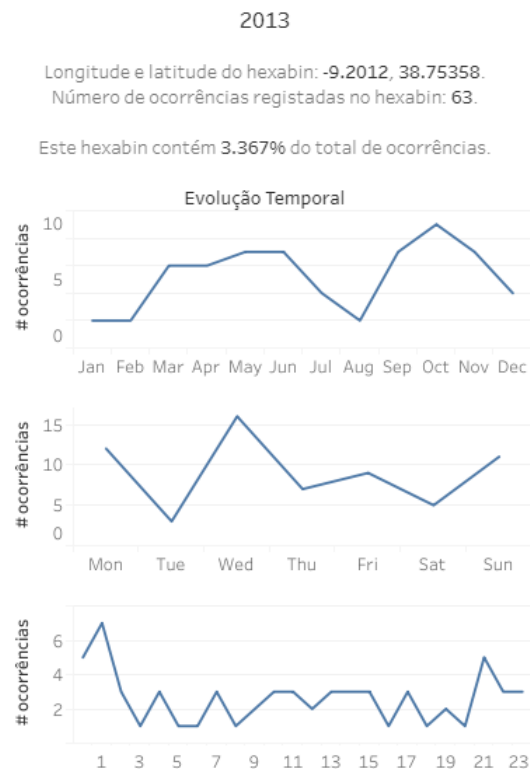


Figura 3.27: Exemplo da *tooltip* T1 do *dashboard* da evolução espacial das tipologias ao longo do tempo.

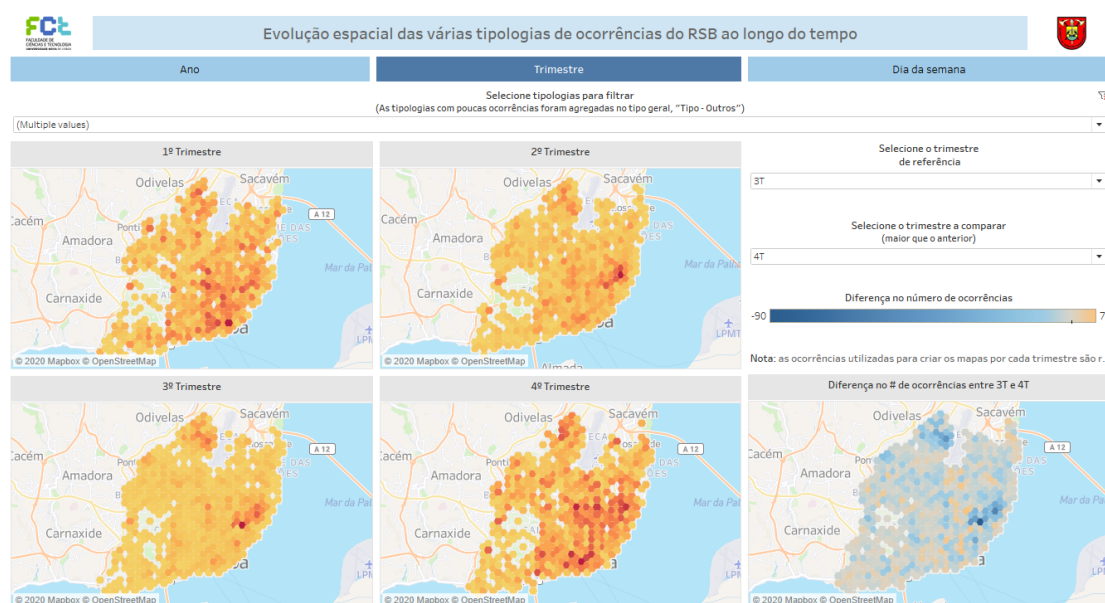


Figura 3.28: *Dashboard* da evolução espacial das tipologias das ocorrências ao longo dos trimestres.



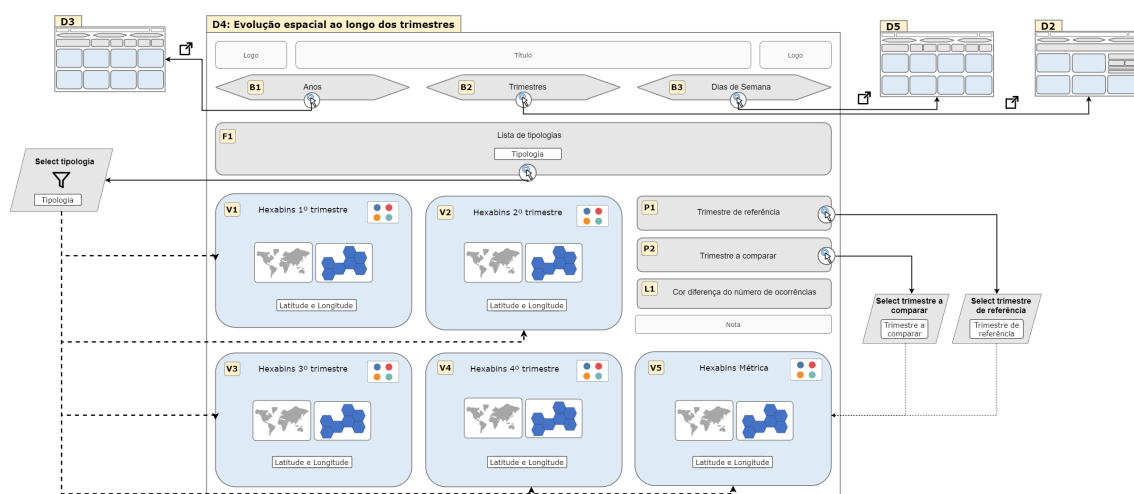


Figura 3.29: Diagrama IVML de interações do *dashboard* da evolução espacial das tipologias das ocorrências ao longo dos trimestres.

acordo com a diferença no número de ocorrências entre dois trimestres. Por fim, existe uma nota no *dashboard* que tem o intuito de alertar o utilizador que neste *dashboard* existe uma dobra temporal. Estamos a lidar com sete anos de dados e por isso a representação de um trimestre é a junção de todos os trimestres presentes nos dados, ou seja, no total sete trimestres.

Quanto às interações presentes neste *dashboard*, as interações com os botões funcionam da mesma forma que no *dashboard* anterior. O botão B1 permite navegar até ao *dashboard* de análise da evolução espacial ao longo dos anos, o botão B2 leva a este próprio *dashboard* e o botão B3 permite navegar até ao *dashboard* de análise ao nível dos dias da semana. O filtro F1 afeta todas as visualizações V1 até V8 através da seleção/remoção de uma ou mais tipologias. A escolha de anos disponíveis nos parâmetros P1 e P2 afeta a visualização de apresenta a diferença entre dois anos, a visualização V8.

Através da figura 3.30 é possível concluir que as *tooltips* presentes neste *dashboard* da granularidade temporal trimestres são idênticas às presentes no *dashboard* que permite analisar a evolução espacial das ocorrências ao longo dos anos. As *tooltips* das visualizações V1 até V4 aparecem no ecrã sempre que se faz *hover* do rato nos hexágonos e apresentam as mesmas informações que a *tooltip* T1 da figura 3.26, sendo que os gráficos de linhas serão referentes às granularidades ano, dia da semana e horas e as informações destes gráficos estão sujeitas às filtrações das tipologias efetuadas no filtro F1. A *tooltip* T2 é iniciada da mesma forma e apresenta informação igual à *tooltip* T2 da figura 3.26.

Por fim, a figura 3.31 apresenta este *dashboard*, mas para a granularidade temporal dias da semana. As figuras 3.32 e 3.33 apresentam os diagramas IVML de interações e *tooltips* do *dashboard* que permite visualizar a existência e movimento de *hotspots* ao longo dos dias da semana, respetivamente.

Como é visível na figura 3.32 este *dashboard* é bastante similar ao *dashboard* que permite fazer a análise a nível de anos. Os botões aqui existentes são iguais aos presentes

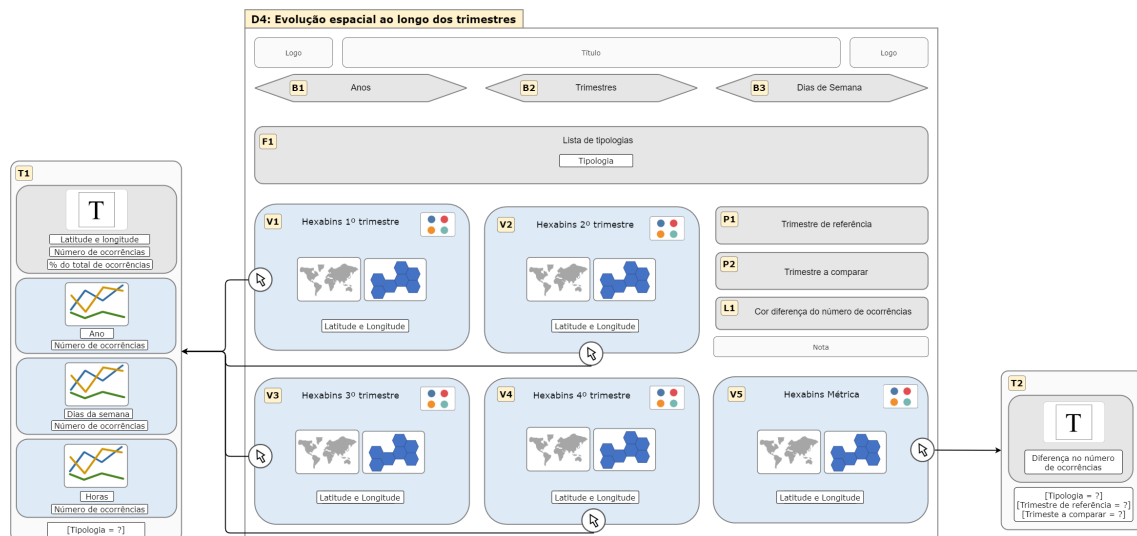


Figura 3.30: Diagrama IVML de tooltips do dashboard da evolução espacial das tipologias das ocorrências ao longo dos trimestres.

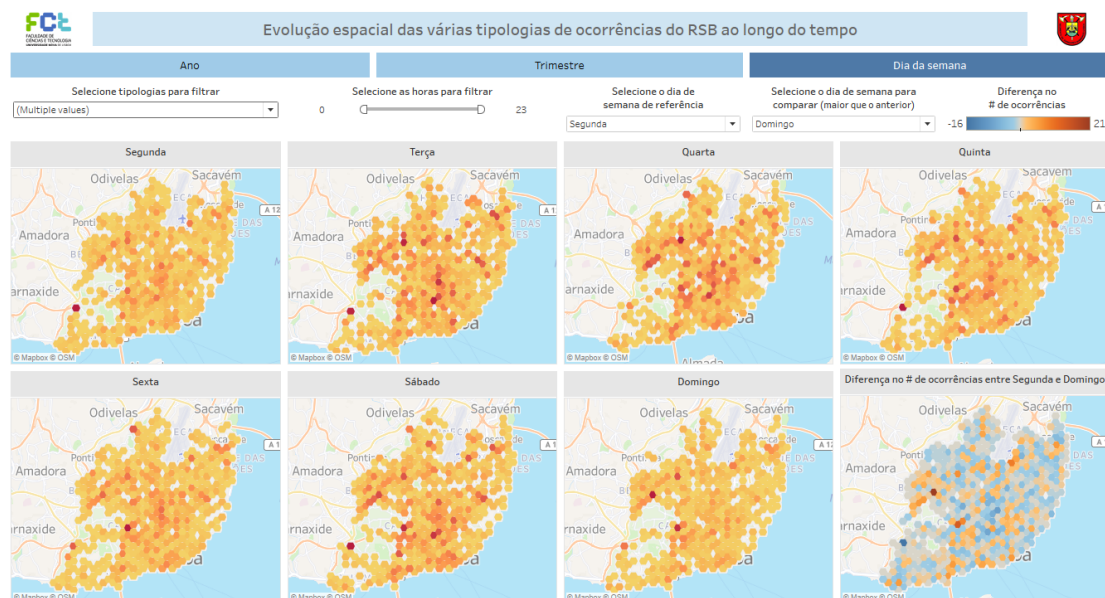


Figura 3.31: Dashboard da evolução espacial das tipologias das ocorrências ao longo dos dias da semana.

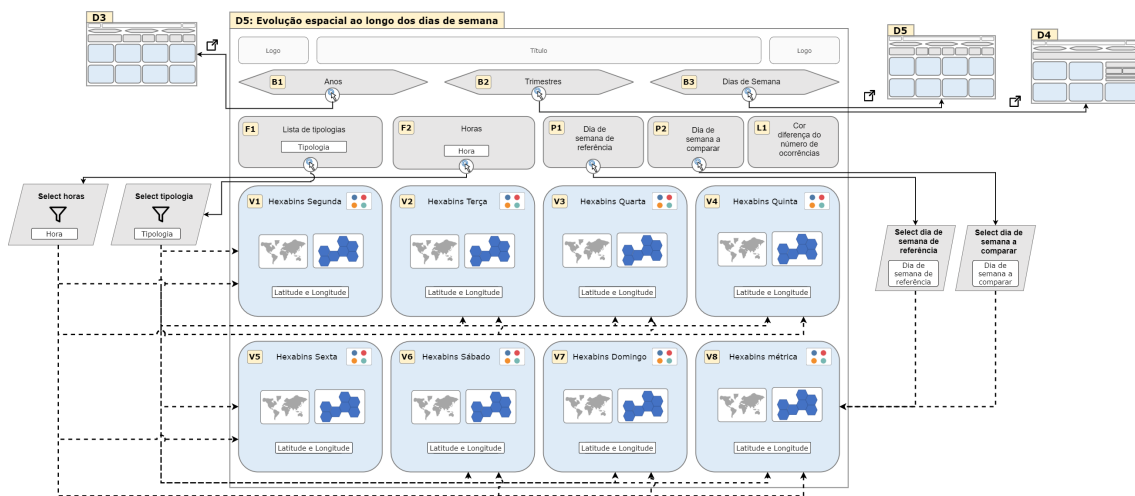


Figura 3.32: Diagrama IVML de interações do *dashboard* da evolução espacial das tipologias das ocorrências ao longo dos dias da semana.

nos outros *dashboards*. O filtro das tipologias funciona da mesma forma que os filtros F1 descritos acima e os parâmetros da escolha de dias da semana para comparação também funcionam da mesma forma que os parâmetros dos outros *dashboards* havendo apenas a diferença de apresentarem os dias da semana e não os anos e trimestres disponíveis no conjunto de dados. Também existem mapas de *hexabins*. Existe um mapa para cada dia da semana, ou seja, existem sete visualizações que utilizam a variável visual cor para colorir os hexágonos de acordo com o número de ocorrências. A visualização V8 é responsável pela comparação entre dias da semana em termos da diferença no número de ocorrências. Esta visualização é um mapa de *hexabins* que utiliza a cor para ver esta diferença. A legenda L1 é a legenda de cor da visualização V8. Este *dashboard* apenas tem um componente a mais que os outros *dashboards*. O filtro F2 é um filtro que apresenta as horas existentes no conjunto de dados. Este filtro é um *slider* que permite selecionar um período de horas.

Relativamente às interações disponibilizadas neste *dashboard* estas são semelhantes às apresentadas para os *dashboards* da análise ao nível de anos e trimestres. Os botões, o filtro para as tipologias e a seleção de parâmetros para controlar a visualização que mostra a diferença entre dois anos funcionam da mesma forma. Este *dashboard* apenas tem uma interação de filtragem que os outros não continham. É possível filtrar os dados que aparecem em todas as visualizações, V1 até V8, por um determinado período de horas. Isto é possível através da seleção desse intervalo de tempo no filtro F2.

As *tooltips* apresentadas na figura 3.33 são idênticas às apresentadas anteriormente. A *tooltip* T1 tem como objetivo apresentar a evolução temporal das ocorrências pertencentes a determinados hexágonos das visualizações V1 até V7. A *tooltip* T2 pertence à visualização V8 e apenas apresenta o valor da diferença no número de ocorrências do hexágono onde o rato está pairando.

Para resumir, com este conjunto de *dashboards* é possível:

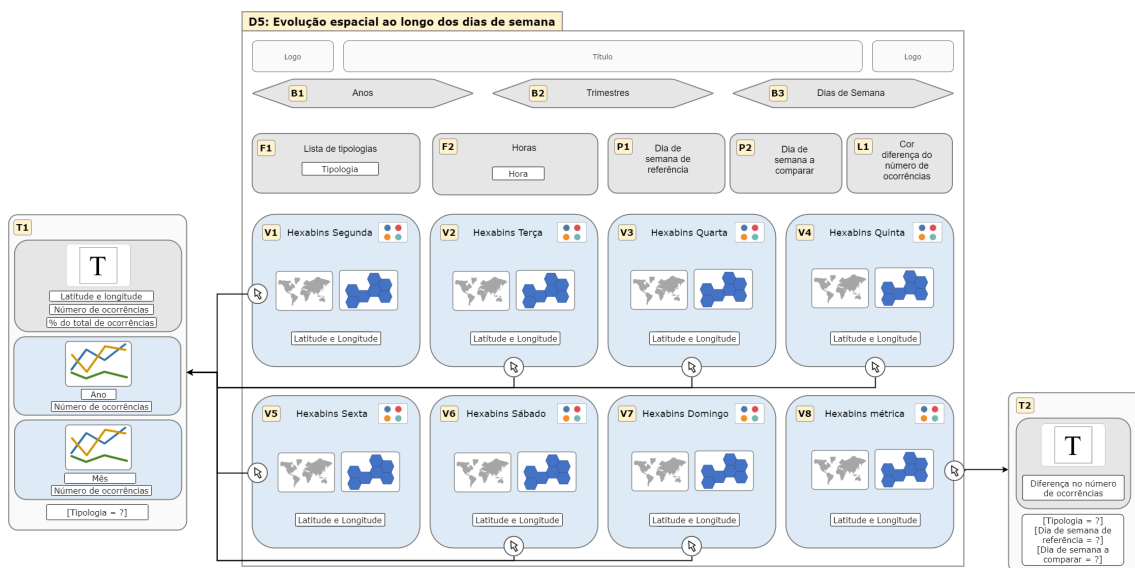


Figura 3.33: Diagrama IVML de *tooltips* do *dashboard* da evolução espacial das tipologias das ocorrências ao longo dos dias da semana.

1. Detetar a existência ou não de *hotspots* ao longo do tempo. Assim é possível responder as questões como "As ocorrências estão dispersas pelo concelho de Lisboa?" ou "Existem *hotspots* de ocorrências?" e "Existe um *hotspot* em todos os anos, trimestres ou dias da semana?";
2. Fazer a análise em diferentes granularidades temporais (anos, trimestres e dias da semana). Isto permite responder às questões acima referidas mas para diferentes granularidades temporais;
3. Analisar se esses *hotspots* se movimentam ao longo do tempo e para onde. Isto permite responder à questão "Os *hotspots* de ocorrências existentes movem-se ao longo do tempo?";
4. Fazer esta análise só para determinadas tipologias de interesse. Assim é possível responder a questões como "Existe alguma zona onde as ocorrências de uma determinada tipologia estão concentradas?";
5. Obter um resumo da evolução entre dois anos que permite comparar o número de ocorrências entre dois anos. Assim o utilizador consegue responder mais facilmente às questões "Quais foram as zonas onde houve um acréscimo no número de ocorrências?" e "Quais foram as zonas onde houve um decréscimo no número de ocorrências?";
6. Obter uma análise temporal das ocorrências de um determinado local. Isto permite responder a questões como "Qual o comportamento temporal desta área?".

A figura 3.34 apresenta quais as visualizações deste *dashboard* que permitem saber os pontos enumerados acima. Cada número presente na figura corresponde ao número

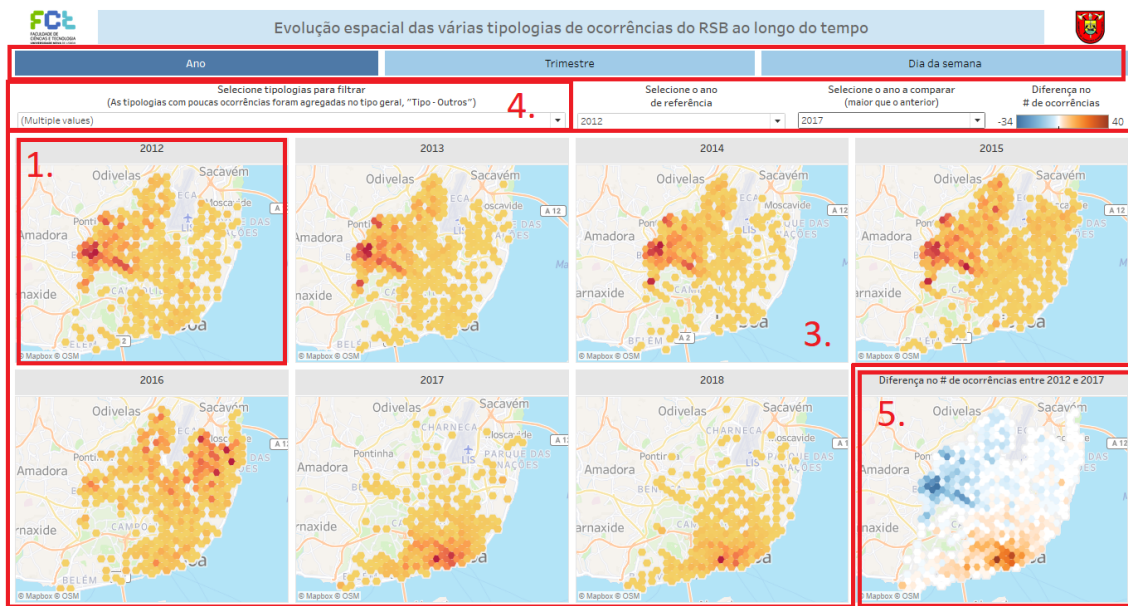


Figura 3.34: Imagem do *dashboard* de evolução espacial das tipologias das ocorrências ao longo dos anos com indicação das visualizações que respondem às questões.

do enumerado. Desta forma é possível fazer um mapeamento entre as visualizações e algumas das questões (enumeradas acima) que estas são capazes de responder. A única visualização que não está acessível na figura 3.34 é a visualização que dá resposta à questão 6. Essa visualização é uma *tooltip* e está presente na figura 3.27.

### 3.5.5 Análise do Comportamento Temporal das Tipologias ao Longo de Várias Granularidades Temporais

Outra questão considerada relevante foi: "Existem tipologias de ocorrências que têm o mesmo comportamento temporal?". Para responder a esta questão foi desenvolvido um *dashboard* que permite fazer uma análise do comportamento temporal das tipologias ao longo do tempo. Para o desenvolvimento deste *dashboard* foi necessário recorrer a mecanismos de *clustering* descritos na seção 3.4. O objetivo do uso deste mecanismo foi obter informações sobre quais as tipologias que partilham do mesmo comportamento temporal. Depois de obter dados sobre qual o comportamento temporal atribuído a cada uma das diferentes tipologias foi necessário criar visualizações que permitissem expor esses dados e adicionar mais algumas análises, nomeadamente, uma análise espacial. Como a definição de comportamento temporal é abrangente, foram definidas diferentes granularidades temporais que seriam alvo de uma análise, as estações do ano, meses, dias da semana, partes do dia e horas. Assim é possível observar quais as tipologias que partilham o mesmo comportamento temporal em diferentes granularidades temporais. Para cada uma dessas granularidades foi desenvolvido um *dashboard*. No entanto, estes *dashboards* são iguais entre si apenas diferindo nos dados de agrupamento utilizados, isto é, apresentam as mesmas visualizações/componentes e interações e apenas diferem na

divisão temporal utilizada. Deste modo, apenas é necessário descrever um deles. Uma imagem de um desses *dashboards* está presente na figura 3.35 e um vídeo demonstrativo das suas funcionalidades está disponível [aqui](#).

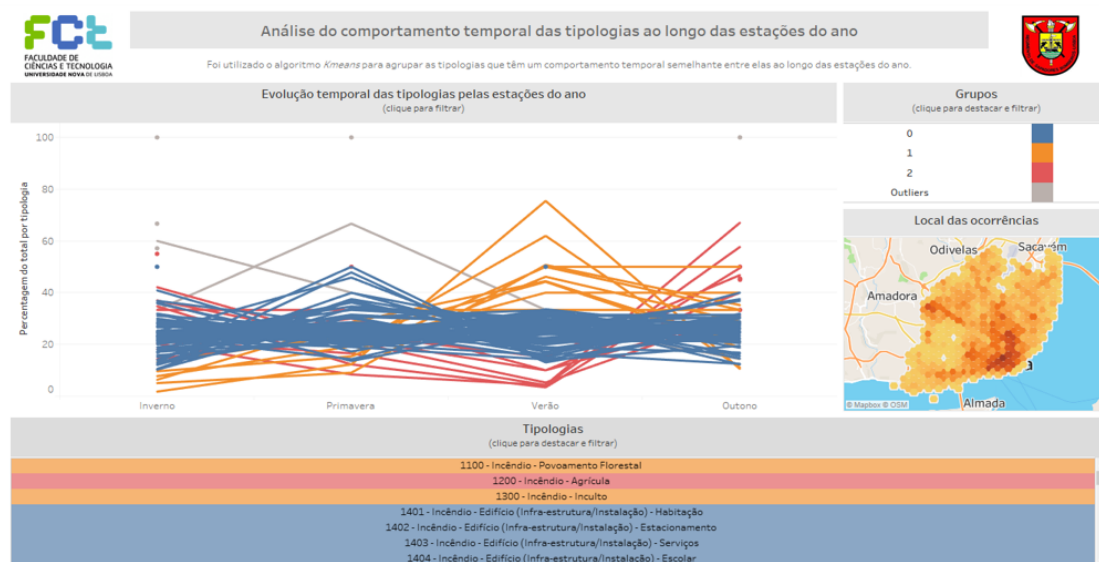


Figura 3.35: Dashboard de análise do comportamento temporal das tipologias ao longo de várias granularidades temporais (estações do ano).

Este *dashboard* foi desenhado para disponibilizar ao utilizador análises relativamente ao comportamento temporal em diversas granularidades temporais, à existência de grupos que partilham esse mesmo comportamento e à distribuição geográfica dos diferentes grupos que partilham do mesmo comportamento temporal.

Para descrever este *dashboard* estão presentes nas figuras 3.36 e 3.37 os diagramas IVML de interações e *tooltips*, respetivamente.

Como ilustrado na figura 3.36 existem 4 visualizações presentes neste *dashboard*. A visualização V1 apresenta a evolução temporal das tipologias ao longo de uma granularidade temporal, isto é, no eixo horizontal está a granularidade temporal e no eixo vertical está a percentagem do total de ocorrências da tipologia. Esta visualização é um gráfico de linhas que contém uma linha para cada tipologia presente nos dados e faz uso da variável visual cor para identificar a que grupo de comportamento temporal a tipologia pertence. Existe uma visualização que funciona como uma legenda de cor no *dashboard*. Esta visualização é a V2 e mostra o mapeamento entre grupos de comportamentos temporais e a cor. A visualização V2 é um gráfico de formas, nomeadamente, quadrados e utiliza a variável visual cor. A visualização V3 permite analisar onde as ocorrências tiveram lugar. Esta visualização é um mapa de *hexabins* que utiliza a cor para mapear o número de ocorrências. Quanto mais escuro for um hexágono mais ocorrências existiram naquele local e quanto mais claro menos ocorrências aconteceram naquele local. Por fim, a visualização V4 apresenta a lista de tipologias existentes nos dados. Esta visualização apenas utiliza texto (a designação da tipologia) e a variável visual cor que, novamente, mapeia para o



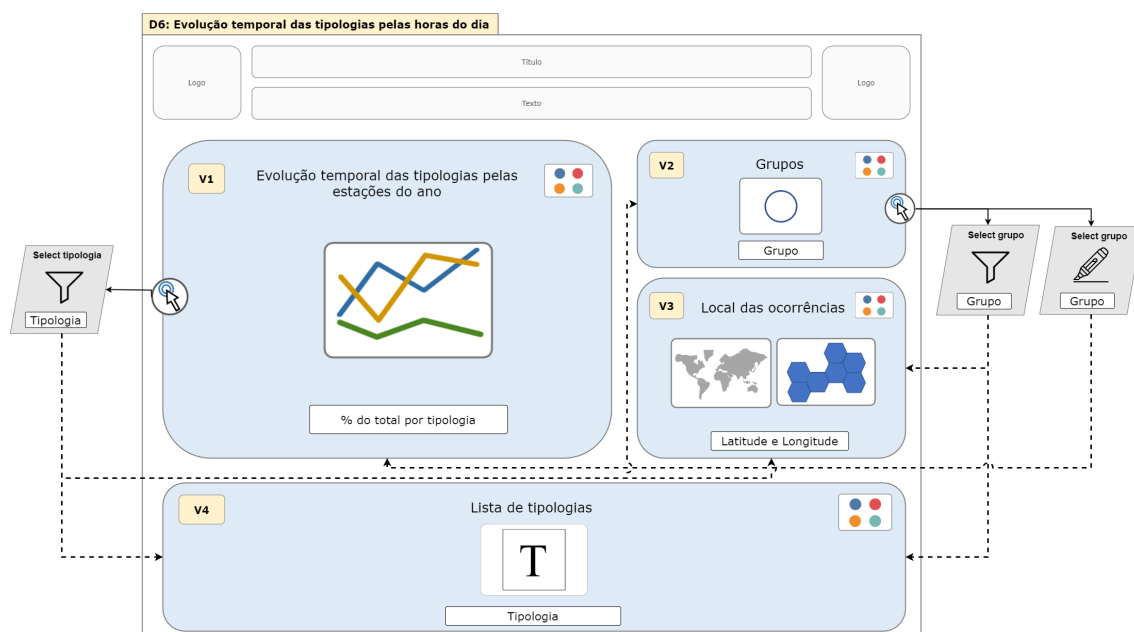


Figura 3.36: Diagrama IVML de interações do *dashboard* de análise do comportamento temporal das tipologias ao longo das estações do ano.

grupo do comportamento temporal e é utilizada para pintar o fundo da célula.

Relativamente às interações presentes neste *dashboard*, existem duas ações de filtragem e uma de destaque. Duas ações estão interligadas, a de destaque e filtragem de um grupo de comportamento temporal. Estas ações acontecem quando se clica na visualização V2 e a ação de destaque apenas afeta a visualização V1, ou seja, no gráfico de linhas são destacadas as linhas das tipologias que apresentam o mesmo comportamento temporal e a ação de filtragem afeta as visualizações V3 e V4, isto é, na V3 apenas serão consideradas para a formação da visualização as localizações das ocorrências que têm a tipologia pertencente ao grupo selecionado e na V4 apenas aparecerão as designações das tipologias que partilham o mesmo comportamento temporal, ficando assim uma lista das tipologias que têm comportamento semelhante. Por fim, é possível analisar as ocorrências de uma só tipologia. Para isto é necessário clicar numa das linhas que representam tipologias na visualização V1 e todas as visualizações do *dashboard* são afetadas. A V2 passa a conter só o grupo do comportamento temporal a que a tipologia pertence, a V3 passa a conter só informação sobre a localização das ocorrências da determinada tipologia e a V4 contém a designação da tipologia selecionada anteriormente.

Seguindo a figura 3.37 podemos concluir que todas as *tooltips* presentes neste *dashboard* são textuais. A *tooltip* T1 aparece no ecrã quando se faz *hover* com o rato num dos pontos temporais de uma das linhas presentes na V1 e é apresentado qual o grupo do comportamento temporal a que a tipologia pertence, o número de ocorrências dessa tipologia naquele ponto temporal e a percentagem do total de ocorrências dessa tipologia naquele ponto do tempo. A *tooltip* T2 é renderizada no ecrã quando se faz *hover* com o rato num dos hexágonos da visualização V3. Esta *tooltip* apresenta as coordenadas geográficas do

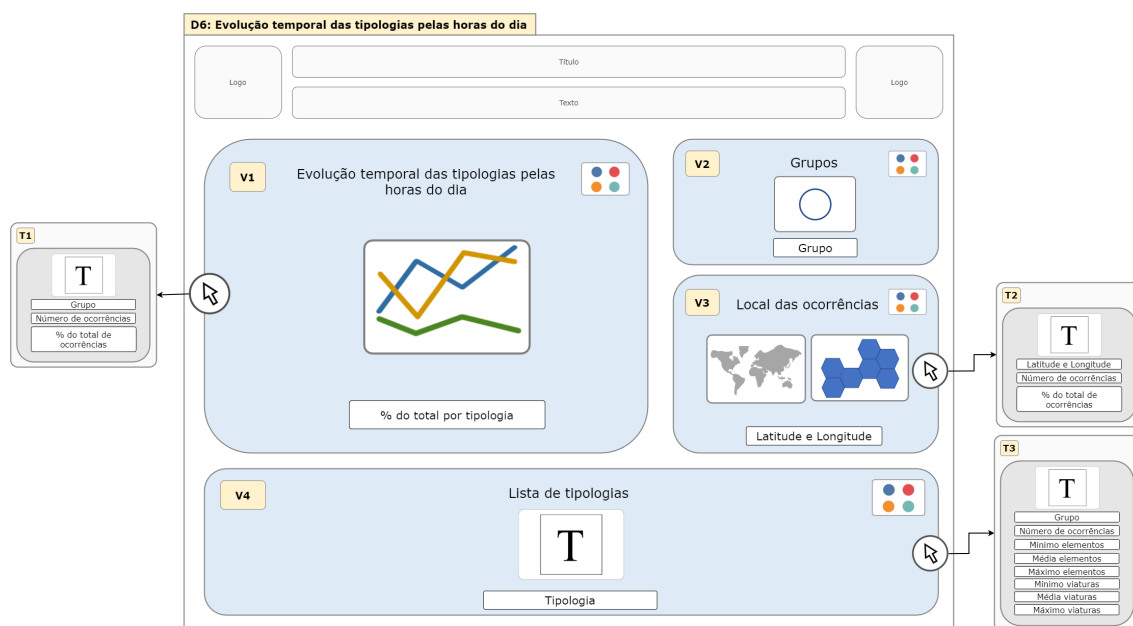


Figura 3.37: Diagrama IVML de *tooltips* do *dashboard* de análise do comportamento temporal das tipologias ao longo das estações do ano.

centro do hexágono, o número de ocorrências que ele contém e a percentagem do total de ocorrências, isto é, a quantidade de ocorrências que o hexágono tem em relação aos outros todos expressa em percentagem. Por fim, a *tooltip* T3 é apresentada quando se faz *hover* numa das designações de tipologias. Esta apresenta informações sobre o grupo a que essa tipologia pertence, o número total de ocorrências dessa tipologia no conjunto de dados e apresenta os valores mínimo, médio e máximo de elementos e viaturas enviadas para o local relativo a todas as ocorrências da tipologia.

Para resumir e concluir, a utilização da ferramenta desenvolvida permite:

1. Analisar qual é o comportamento das tipologias ao longo do tempo em diferentes granularidades temporais. Isto possibilita a obtenção de respostas a questões como "Existe sazonalidade nas tipologias?" e "Qual o comportamento temporal da tipologia X ao longo das estações do ano, meses, dias da semana, partes do dia e horas?";
2. Saber quais as tipologias que partilham o mesmo comportamento temporal e qual é esse comportamento. Isto permite obter respostas a questões como "Quais são as tipologias que partilham o mesmo comportamento temporal?", "A tipologia X e Y têm um comportamento temporal semelhante?" e "Quais são as tipologias que apresentam sazonalidade no Verão?";
3. Ter uma visão geral da distribuição geográfica das ocorrências pelo concelho de Lisboa. Isto possibilita responder a questões como "As ocorrências das tipologias que apresentam sazonalidade no Verão estão concentradas em alguma zona de



Lisboa?"e "As ocorrências que não partilham o mesmo comportamento temporal partilham da mesma distribuição geográfica?"

4. Ter visão geral de meios mobilizados em diferentes tipologias. Isto permite responder a questões como "Em média, quantos elementos e viaturas são mobilizados pela tipologia que apresenta sazonalidade no Verão?".

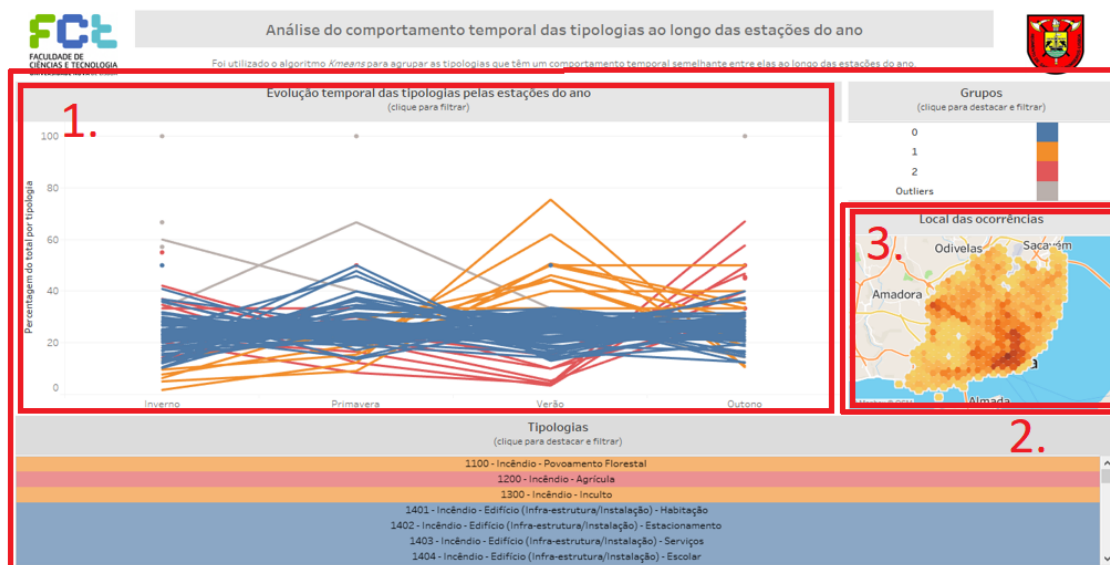


Figura 3.38: Imagem do *dashboard* da análise do comportamento temporal das tipologias ao longo das estações do ano com indicação das visualizações que respondem às questões.

A figura 3.38 apresenta quais as visualizações deste *dashboard* que permitem saber os pontos enumerados acima. Cada número presente na figura corresponde ao número do enumerado. Desta forma é possível fazer um mapeamento entre as visualizações e algumas das questões (enumeradas acima) que estas são capazes de responder. A única visualização que serve para obter resposta a uma questão que não está instanciada nesta figura é a questão 4. A visualização está presente na *tooltip* T3.

### 3.5.6 Análise do Comportamento Espacial das Ocorrências

Após ser desenvolvido um *dashboard* para analisar o comportamento temporal das tipologias das ocorrências foi desenvolvido outro *dashboard* que permite analisar o comportamento espacial das ocorrências do RSB. A questão relevante a que este *dashboard* pretende responder é: "Existem agrupamentos espaciais de ocorrências no concelho de Lisboa?". Tal como no *dashboard* anterior, para responder a esta questão foi necessário recorrer a algoritmos de agrupamento para definir quais são os grupos espaciais das ocorrências existentes. O processo de agrupamento espacial foi descrito na seção 3.4. Tal como anteriormente referido na seção 3.4, foram feitos três agrupamentos, um com todas as ocorrências, outro com todas as ocorrências menos alguns "Serviços" e outro só com as

ocorrências de emergência. Para cada um desses agrupamentos foi desenvolvido um *dashboard*. Estes *dashboards* apresentam todos as mesmas componentes e interações, sendo que apenas diferem nos dados de agrupamentos utilizados. Portanto, para efeitos de explicação apenas é necessário descrever um destes *dashboards*. A imagem do *dashboard* que apresenta os resultados do agrupamento com todas as ocorrências está presente na figura 3.39 e um vídeo demonstrativo das suas funcionalidades está disponível [aqui](#).

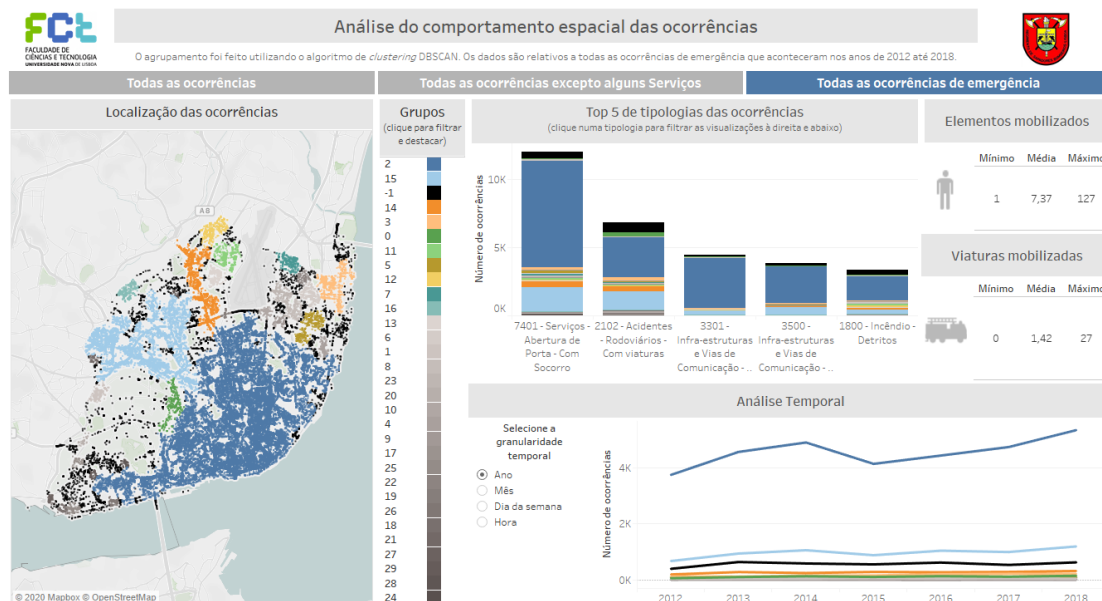


Figura 3.39: Dashboard de análise do comportamento espacial das ocorrências.

Este *dashboard* foi desenhado para disponibilizar ao utilizador análises e comparações, dos diferentes grupos espaciais resultantes do agrupamento, para a distribuição das ocorrências por tipologias em que o utilizador esteja interessado, aos meios humanos e viaturas mobilizadas por ocorrências e comportamentos temporais a diferentes granularidades temporais, ao nível anual, mensal, dia da semana e horário.

Para descrever o *dashboard* estão presentes nas figuras 3.40 e 3.41 os diagramas IVML de interações e *tooltips*, respetivamente.

Como ilustrado na figura 3.40 existem três botões, seis visualizações e um parâmetro. Os botões presentes neste *dashboard* servem para navegar entre os outros *dashboards* com os diferentes tipos de agrupamentos. O único parâmetro neste *dashboard* é o parâmetro P1. Este serve para selecionar a granularidade temporal para a análise da evolução temporal das ocorrências. Existem quatro valores possíveis para este parâmetro: "Ano", "Mês", "Dia da semana" e "Hora". Relativamente às visualizações, a visualização V1 é um mapa que apresenta, em forma de pontos, todas as ocorrências presente no conjunto de dados de acordo com as suas coordenadas geográficas. Esta visualização utiliza a variável visual cor para identificar quais os grupos (resultantes do *clustering*) atribuídos a cada ocorrência. A coloração deste mapa foi descrita na seção 3.4, onde a cor preta representa ruído, cores coloridas representam ocorrências de grupos de dimensões maiores e cores

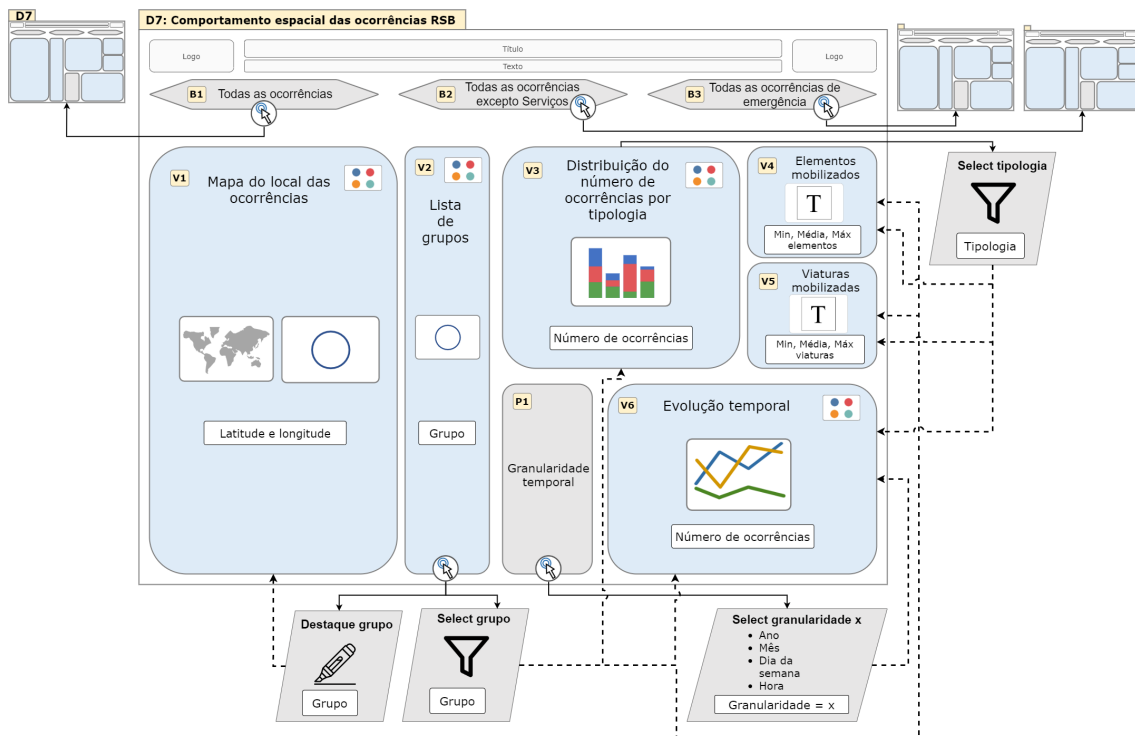


Figura 3.40: Diagrama IVML de interações do *dashboard* da evolução espacial das tipologias das ocorrências ao longo do tempo.

numa escala de cinzentos representam ocorrências de grupos de dimensões menores. A visualização V2 apresenta a lista de grupos resultantes do processo de agrupamento. Esta visualização apresenta o identificador de cada grupo e um quadrado que está colorido de acordo com uma cor que irá identificar o grupo em todo o *dashboard*. A legenda presente nesta visualização está ordenada de acordo com o número de ocorrências presentes em cada grupo, ou seja, os grupos maiores estão em cima e grupos menores estão abaixo. A visualização V3 apresenta a distribuição do número de ocorrências por tipologia. Sendo que existem muitas tipologias, esta distribuição apenas à visível para o *top* cinco das tipologias com mais registos. Esta visualização é um gráfico de barras e tem uma barra para cada tipologia presente no *top* cinco. Cada barra desta visualização está subdividida de acordo com o número de ocorrências que cada grupo tem. Para identificar estes segmentos presentes em cada barra é utilizada a cor que identifica do grupo a que o segmento se refere. A visualização V4 e V5 apresentam os valores do mínimo, média e máximo do número de elementos e viaturas mobilizadas em forma textual. Por fim, a visualização V6 é um gráfico de linhas que contém uma linha para cada grupo resultante do agrupamento e, para distinguir a qual grupo a linha se refere, foi utilizada a variável visual cor. Esta visualização tem no eixo horizontal uma determinada granularidade temporal, anos, meses, dias da semana ou horas, e no eixo vertical o número de ocorrências. Tem como objetivo demonstrar a evolução temporal dos diferentes grupos.

Quanto às interações, este *dashboard* disponibiliza ao utilizar três interações de navegação, duas de filtragem, uma de destaque e uma de seleção de parâmetros. As três interações de navegação podem ser acionadas quando se clica num dos três botões, B1, B2 e B3. Ao se clicar no botão B1 regressa-se ao *dashboard* que tem o agrupamento de todas as ocorrências presentes no conjunto de dados, ao se clicar no B2 o utilizador é redirecionado para o *dashboard* que tem o agrupamento de todas as ocorrências excepto alguns "Serviços" e, por fim, ao se clicar no B3 o utilizador é redirecionado para o *dashboard* que apresenta os resultados do agrupamento das ocorrências de emergência. A ação de seleção de parâmetros é acionada quando se escolhe um dos valores disponíveis no parâmetro P1. Esta ação afeta apenas a visualização V6. A visualização V6 passará a apresentar a evolução do número de ocorrências ao longo da granularidade temporal selecionada. Quanto às ações de filtragem, é possível filtrar as visualizações V4, V5 e V6 para apenas apresentarem dados sobre uma determinada tipologia. Isto é possível através de um clique na visualização V4 que apresenta a distribuição do número de ocorrências por tipologia. A outra interação de filtragem está ligada à ação de destaque. Ambas se iniciam quando se seleciona um ou mais grupos presentes na lista de grupos na visualização V2. A ação de destaque afeta apenas a visualização V1 e é feito um destaque no mapa das ocorrências que pertencem ao(s) grupo(s) selecionado(s). A ação de filtragem afeta todas as outras visualizações, da V3 à V6. Nestas visualizações apenas aparecerão informações relativas ao(s) grupo(s) selecionado(s). Desta forma é possível selecionar dois ou mais grupos para comparar as suas características.

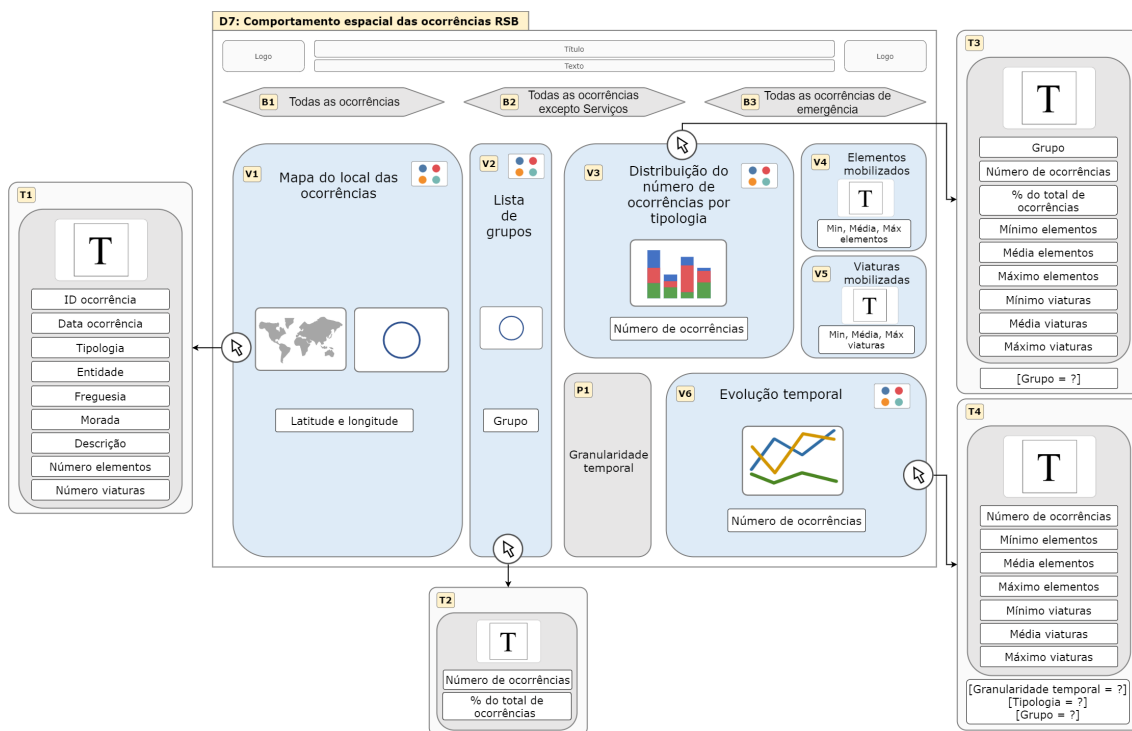


Figura 3.41: Diagrama IVML de tooltips do dashboard da evolução espacial das tipologias das ocorrências ao longo do tempo.

Segundo a figura 3.41, que apresenta as *tooltips*, as informações dadas pelas *tooltips* são todas em forma textual. A *tooltip* T1 é referente à visualização V1. Esta aparece no ecrã quando se faz *hover* do rato num ponto do mapa, uma ocorrência. A *tooltip* apresenta as seguintes informações sobre a ocorrência: o seu identificador, a data e hora, a tipologia, a entidade que iniciou a ocorrência, a freguesia, a morada, a descrição e o número de meios mobilizados para o local. A *tooltip* T2 fica disponível quando se faz *hover* do rato num dos grupos presentes na lista da visualização V2 e são apresentadas as seguintes informações sobre o grupo: número de ocorrências presentes no grupo e a percentagem do total de ocorrências no conjunto de dados que o grupo contém. A *tooltip* T3 aparece no ecrã quando se faz *hover* num dos segmentos presentes nas barras da visualização V3. Esta apresenta as seguintes informações sobre as ocorrências da determinada tipologia no determinado grupo: número de ocorrências, percentagem do total de ocorrências, valores mínimo, médio e máximo tanto do número de elementos mobilizados como de viaturas. Por fim, a *tooltip* T4 é relativa à visualização V6 e é disponibilizada quando se faz *hover* num ponto do tempo numa das linhas que representam a evolução temporal de um grupo. Esta apresenta as seguintes informações sobre as ocorrências daquele grupo naquele ponto do tempo: número de ocorrências e os valores mínimo, médio e máximo do número de elementos e viaturas mobilizados para o local.

Para concluir e resumir, este *dashboard* permite o utilizador:

1. Analisar quais são os grupo espaciais das ocorrências do [RSB](#) no concelho de Lisboa. Isto possibilita responder a questões como "Quais os grupos espaciais de ocorrências?" e "Quais são os maiores grupos e onde se localizam?";
2. Saber quais são as características de cada um desses grupos. As características podem ser analisadas em termos de tipologias, meios mobilizados para o local das ocorrências (Homens e viaturas) e comportamento temporal do grupo espacial. Isto permite obter respostas a questões como "Qual o comportamento temporal do grupo espacial X?" e "Qual a tipologia das ocorrências mais comum no grupo X?";
3. Comparar grupos através das suas características. Isto possibilita saber a resposta a perguntas como "O grupo X e o grupo Y partilham do mesmo comportamento temporal?" e "O grupo X e o grupo Y têm a mesma distribuição por tipologias?";
4. Comparar grupos através das suas características utilizando apenas dados de tipologias alvo de interesse de análise. Esta funcionalidade do *dashboard* permite responder a questões como "Como as ocorrências da mesma tipologia se comportam temporalmente nos diferentes grupos?";
5. Fazer todo o que está acima descrito, mas para diferentes tipos de agrupamentos, considerando todas as ocorrências no conjunto de dados, todas as ocorrências excepto algumas do tipo "Serviços" e apenas ocorrências de emergência. A resposta a

questões como "Qual é o comportamento temporal do grupo X ao longo dos anos, meses, dias da semana e horas?" são possíveis de obter com este *dashboard*.

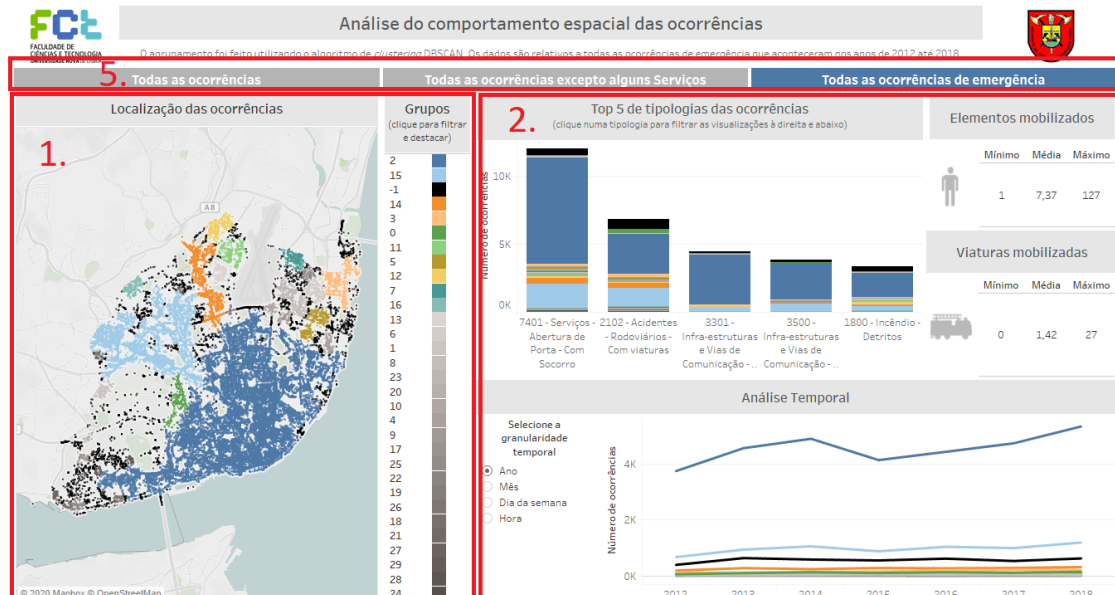


Figura 3.42: Imagem do *dashboard* da análise do comportamento espacial das ocorrências do RSB com indicação das visualizações que respondem às questões 1, 2 e 5.

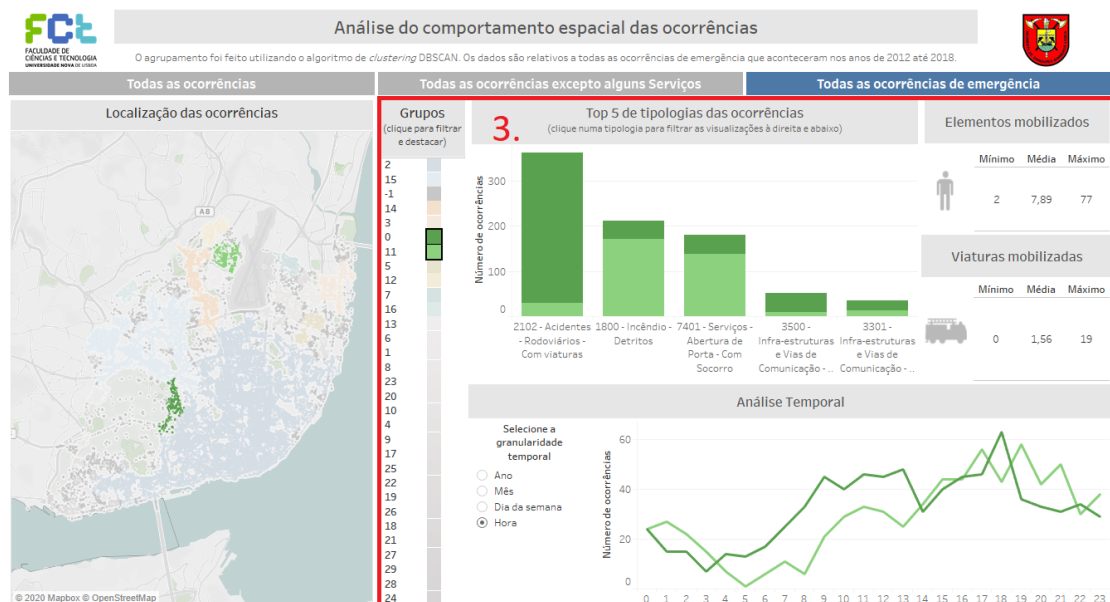


Figura 3.43: Imagem do *dashboard* da análise do comportamento espacial das ocorrências do RSB com indicação das visualizações que respondem à questão 3.

As figuras 3.42, 3.43 e 3.44 apresentam quais as visualizações deste *dashboard* que permitem saber os pontos enumerados acima. Cada número presente na figura corresponde ao número do enumerado. Desta forma é possível fazer um mapeamento entre



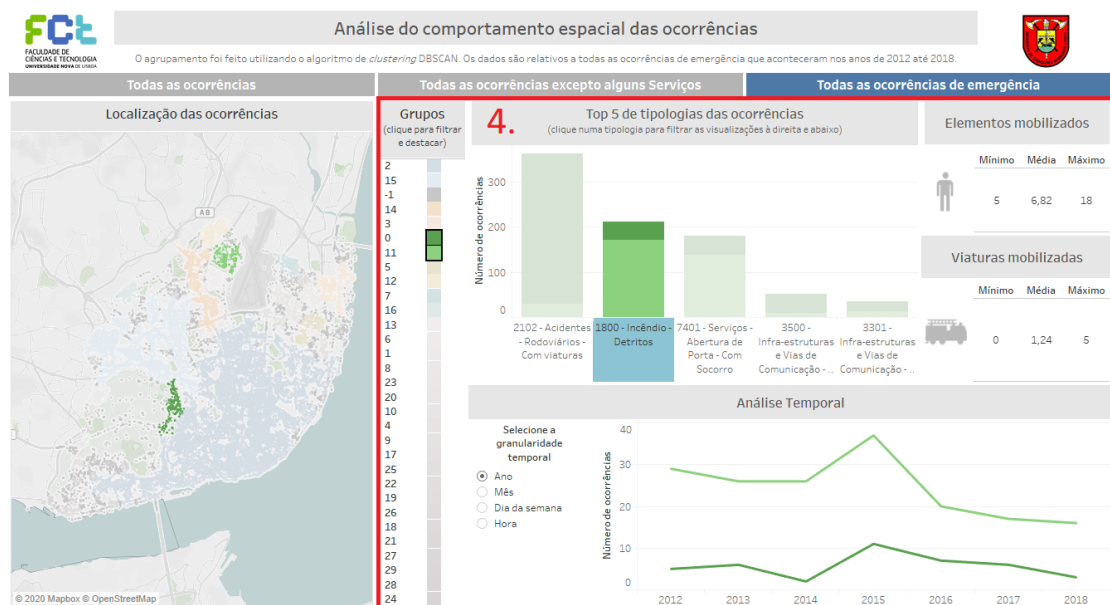


Figura 3.44: Imagem do *dashboard* da análise do comportamento espacial das ocorrências do RSB com indicação das visualizações que respondem à questão 4.

as visualizações e algumas das questões (enumeradas acima) que estas são capazes de responder. Este *dashboard* necessitou de mais figuras para representar onde se pode obter respostas às perguntas enumeradas, pois algumas das respostas só podem ser obtidas após a realização de interações.

### 3.5.7 Análise das Ocorrências em Redor de Pontos de Interesse em Lisboa

Com a apresentação dos *dashboards* anteriores ainda faltam responder a algumas questões relevantes. Uma dessas questões é: "O que acontece em termos de ocorrências em redor de pontos de interesse do concelho de Lisboa?". Para responder a tal pergunta foi desenvolvido um *dashboard* que permite fazer a análise das ocorrências que acontecem em redor de pontos de interesse de Lisboa. Para o desenvolvimento deste *dashboard* houve a necessidade de ser criado um novo conjunto de dados. Este conjunto de dados irá conter alguns dos pontos de interesse existentes no concelho de Lisboa. Os pontos de interesse foram divididos em categorias e essas categorias podem ter subcategorias. Essas categorias e subcategorias podem ser:

- Transportes (metro Lisboa, GIRA, estações ferroviárias e estações fluviais);
- Ensino (ensino superior e ensino secundário );
- Saúde (centros de saúde, hospitais e lares);
- Desporto (estádios de futebol);
- Cultura (teatro, locais de eventos e museus);

## CAPÍTULO 3. TRABALHO REALIZADO

- Lazer (centros comerciais, parques e jardins, cinemas, casinos, terminais de cruzeiros, marinas e docas, elevadores/ascensores e hotéis).

Os atributos deste novo conjunto de dados estão presentes na tabela 3.15. Relativamente ao raio do *buffer*, estes estão definidos de acordo com a subcategoria dos pontos de interesse. Atualmente existem dois valores possíveis: 250 ou 400 metros de raio. Por exemplo, hospitais são edifícios de maiores dimensões, por isso, está atribuído o valor do raio maior, 400 metros. Os dados, relativamente ao nome e localização, foram retirados maioritariamente do *site* Lisboa Aberta. A criação deste conjunto de dados foi feita em colaboração com outros colegas de dissertação que também tinham interesse em obter pontos de interesse do concelho de Lisboa.

Tabela 3.15: Atributos do conjunto de dados dos pontos de interesse.

Atributo	Descrição
Categoria	Categoria a que o ponto de interesse pertence
Subcategoria	Subcategoria a que o ponto de interesse pertence
Nome	Nome do ponto de interesse
Latitude	Latitude do ponto de interesse
Longitude	Longitude do ponto de interesse
Tamanho <i>buffer</i>	Tamanho do raio do <i>buffer</i> a considerar para a análise das ocorrências em redor do ponto de interesse (em metros)

Uma imagem do *dashboard* desenvolvido e que utiliza este novo conjunto de dados está presente na figura 3.45 e um vídeo demonstrativo das suas funcionalidades está disponível [aqui](#).

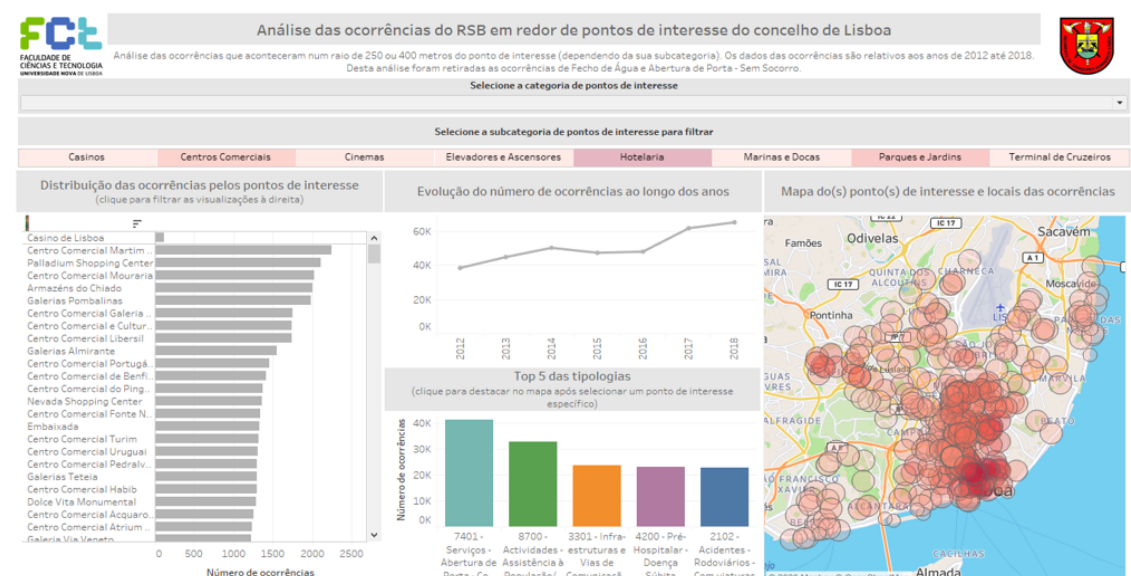


Figura 3.45: Dashboard de análise das ocorrências em redor de pontos de interesse no concelho de Lisboa.



Para descrever os componentes e as funcionalidades deste *dashboard* estão presentes nas figuras 3.40 e 3.41 os diagramas IVML de interações e *tooltips*, respetivamente.

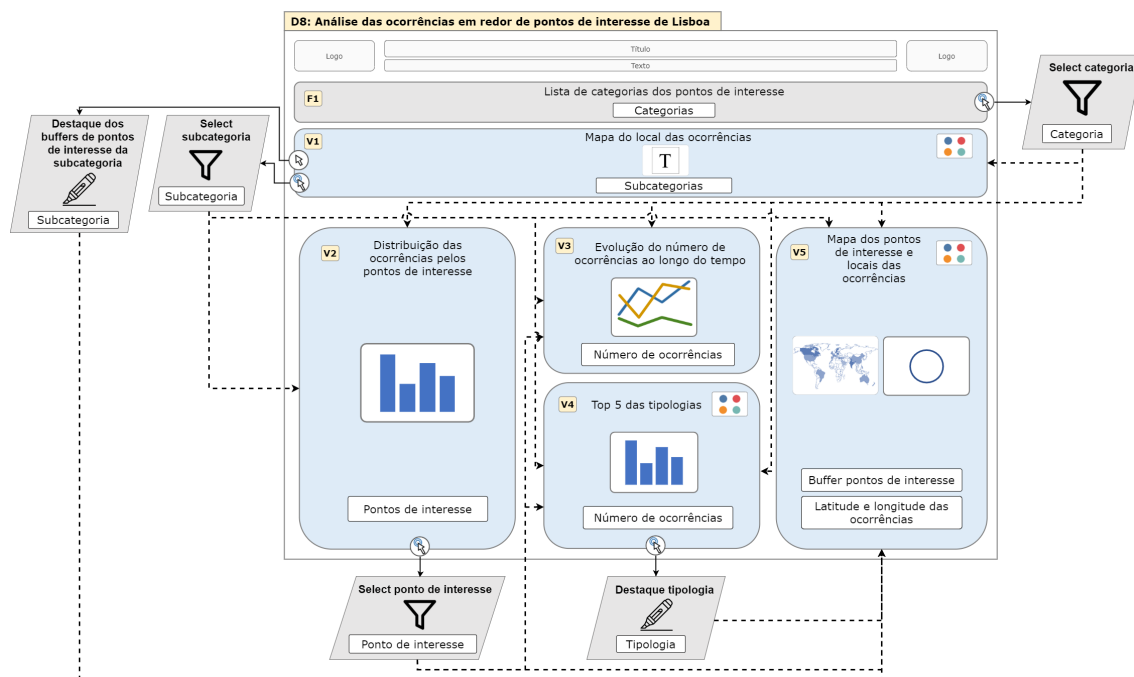


Figura 3.46: Diagrama IVML de interações do *dashboard* de análise das ocorrências em redor de pontos de interesse no concelho de Lisboa.

Como mostra a figura 3.46, este *dashboard* é composto por um filtro e cinco visualizações. O único componente filtro que existe é o filtro F1. Este filtro apresenta uma lista das categorias dos pontos de interesse que existem no novo conjunto de dados anteriormente referido. A visualização V1 apresenta, de forma textual, quais são as subcategorias existentes nas categorias dos pontos de interesse. Esta visualização faz ainda uso da cor para colorir os fundos dos textos das subcategorias de acordo com o número de ocorrências que nelas estão presentes. A visualização V2 apresenta a distribuição do número de ocorrências pelos vários pontos de interesse. Para tal é utilizado um gráfico de barras. A visualização V3 é um simples gráfico de linhas que demonstra a evolução do número de ocorrências ao longo dos anos. A visualização V4 apresenta a distribuição do número de ocorrências pelas tipologias. Novamente, dado o elevado número de tipologias, esta análise foi restringida ao *top* cinco das tipologias com mais registos de ocorrências. Esta visualização é um gráfico de barra que utiliza a variável visual cor para distinguir as tipologias, ou seja, cada barra tem uma única cor que a distingue das outras barras. Por fim, a visualização V5 permite visualizar no mapa os *buffers* dos pontos de interesse. Estes *buffers* são círculos, cujo centro é o local do ponto de interesse. O raio do círculo atualmente está definido no conjunto de dados, sendo que, sempre que a análise o peça, poderá ser alterado e redefinido. A visualização V5 faz ainda uso da cor para o número de ocorrências que o *buffer* contém, ou seja, quanto mais escuro o círculo mais ocorrências ele tem.

Relativamente às interações disponibilizadas por este *dashboard* aos utilizadores, existem três ações de filtragem e duas de destaque. A primeira ação de filtragem acontece quando se seleciona um valor da lista das categorias dos pontos de interesse, presente no filtro F1. A escolha de um valor desta lista vai afetar todo o *dashboard*, ou seja, todas as visualizações, V1 até V5. As visualizações afetadas apenas irão utilizar as ocorrências que ficam dentro dos *buffers* dos pontos de interesse da categoria selecionada. A visualização V1 disponibiliza duas ações diferentes, uma ação de filtragem por subcategoria e uma ação de destaque de pontos de interesse que partilham a subcategoria. Relativamente à ação de filtragem, esta inicia-se com um clique na subcategoria de interesse presente na visualização V1 e afeta as visualizações de V2 até V5. Estas apenas passam a apresentar pontos de interesse da subcategoria selecionada e dados de ocorrências que estão contidas nos *buffers* desses pontos de interesse. A ação de destaque inicia-se com um *hover* do rato sobre as subcategorias da visualização V1 e afeta apenas a visualização que apresenta o mapa com os *buffers*. O mapa passa apenas a apresentar os *buffers* dos pontos de interesse da subcategoria selecionada. A última ação de filtragem é iniciada na visualização V2 com a seleção de um ponto de interesse. Esta ação irá afetar as visualizações V3 à V5, sendo que, estas passarão apenas a mostrar dados relativos àquele ponto de interesse e às ocorrências que fazem parte do seu *buffer*. Esta ação tem um efeito especial na visualização V5. É feito um *zoom in* no mapa para apenas ser mostrado o *buffer* do ponto de interesse em causa e dentro do *buffer* irão aparecer as suas ocorrências, em forma de pontos, de acordo com a localização das mesmas. A utilização da variável cor desta visualização é mudada para colorir os pontos das ocorrências do mapa de acordo com a sua tipologia. Se a tipologia da ocorrência estiver no *top* cinco é atribuído ao ponto a cor que a identifica na visualização V4, se não estiver no *top* é atribuída a cor cinzenta. A figura 3.47 evidencia o resultado desta ação de filtragem no *dashboard*. Por fim, a última ação de destaque inicia-se quando se clica numa tipologia presente na visualização V4. Esta ação apenas afeta o mapa da visualização V5 e são destacadas as ocorrências que têm a tipologia que foi selecionada.

Seguindo a figura 3.48, existem dois tipos de *tooltips*, as textuais e as que contêm gráficos. A *tooltip* T1 é textual e pertence à visualização V1. É acessível através de um *hover* de uma subcategoria presente em V1 e apresenta as seguintes informações sobre as ocorrências que fazem parte dos *buffers* dos pontos de interesse com essa categoria: percentagem do total de ocorrências, valores mínimo, médio e máximo de elementos e viaturas mobilizadas. A *tooltip* T2 é textual e aparece no ecrã quando se faz *hover* numa das barras que representam pontos de interesse na visualização V2 ou numa das barras que representam uma tipologia em V4. Esta apresenta as seguintes informações sobre as ocorrências em redor do ponto de interesse ou da tipologia: número de ocorrências, valores mínimo, médio e máximo de elementos e viaturas mobilizadas para o local. A *tooltip* T4 é também textual e aparece sempre que se faz *hover* num *buffer* presente em V5. As seguintes informações são apresentadas nesta *tooltip*: categoria e subcategoria do ponto de interesse, raio do *buffer* e número de ocorrências presentes no *buffer*. A visualização

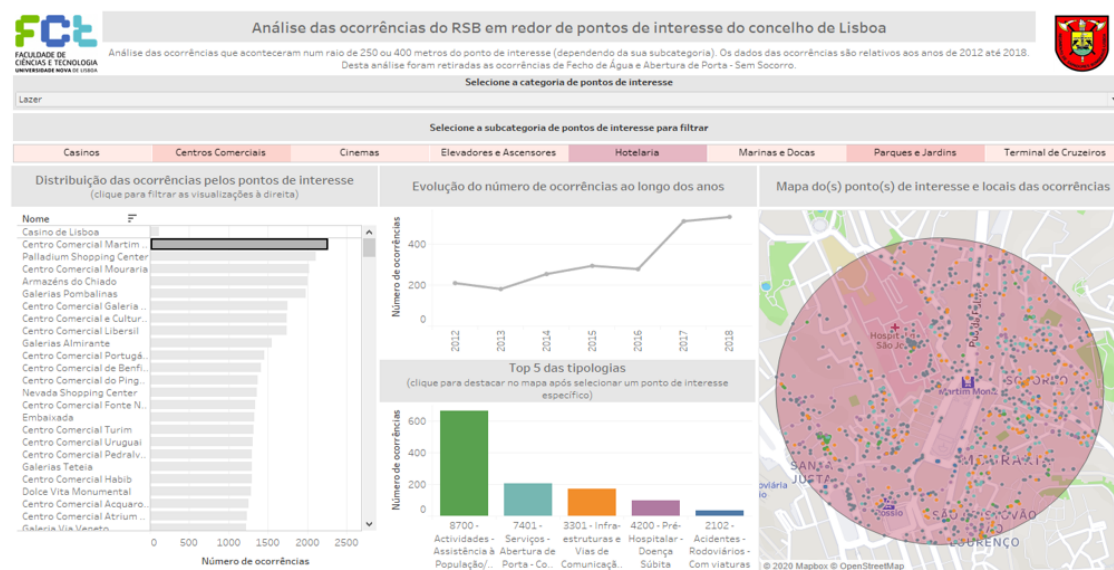


Figura 3.47: Dashboard de análise das ocorrências em redor de pontos de interesse no concelho de Lisboa após interação de escolha de um ponto de interesse.

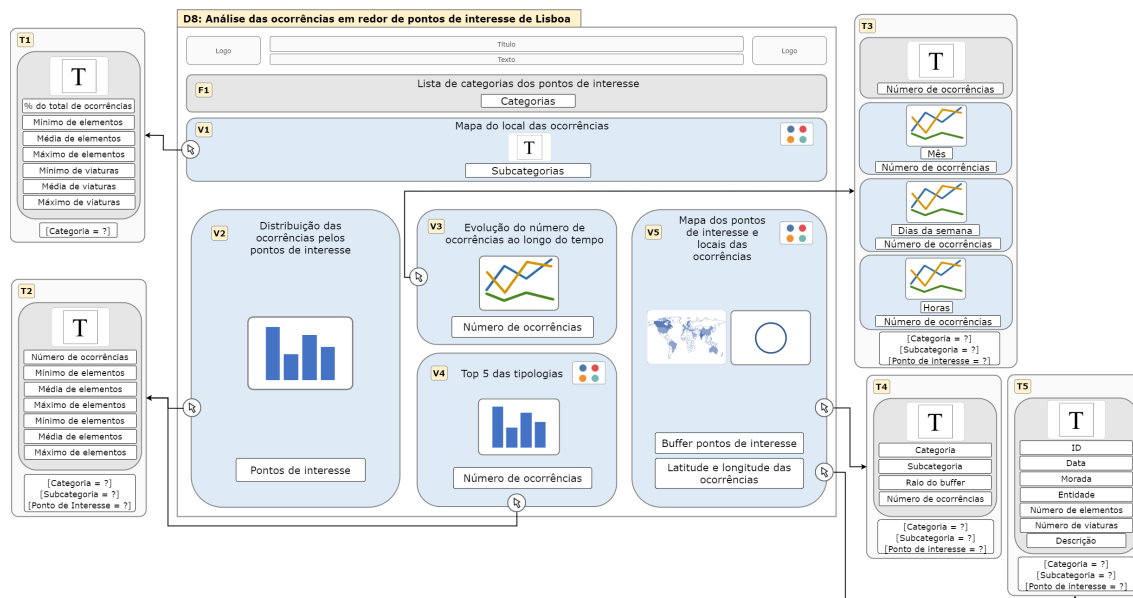


Figura 3.48: Diagrama IVML de tooltips do dashboard de análise das ocorrências em redor de pontos de interesse no concelho de Lisboa.

V5, quando se faz *zoom in* no mapa através da interação de filtragem apresenta ainda outra *tooltip*. Essa *tooltip* é a T5 que aparece quando se faz *hover* num ponto que representa uma ocorrência. As seguintes informações serão apresentadas sobre a ocorrência: o identificador, a data e hora, a morada, a entidade que iniciou a ocorrência, o número de meios mobilizados para o local e a sua descrição. Por fim, a *tooltip* T3 é uma *tooltip* que apresenta gráficos que permitem ver a evolução temporal do número de ocorrências em termos de meses, dias da semana e horas. Esta *tooltip* é disponibilizada quando se faz *hover* do rato num ponto do tempo da visualização V3. A *tooltip* é semelhante à apresentada visualmente na figura 3.27.

Para concluir e resumir, este *dashboard* permite o utilizador:

1. Observar onde estão os pontos de interesse no concelho de Lisboa. Esta funcionalidade permite o utilizador obter respostas a questões como "Onde se localizam os pontos de interesse?", "Onde existe maior concentração de pontos de interesse?" e "Quais os pontos de interesse que têm maior número de ocorrências?";
2. Observar e analisar os pontos de interesse de acordo com a sua categoria e/ou subcategoria. Isto permite responder às questões acima referidas mas apenas para pontos de interesse com determinadas características de uso. Uma questão que pode ser respondida com esta funcionalidade é "Onde se concentram mais os pontos de interesse da categoria de Lazer?";
3. Analisar o que acontece em redor desses pontos de interesse em termos de evolução temporal, tipologias e distribuição geográfica das ocorrências e meios mobilizados. Isto possibilita obter respostas a questões como "Qual o ponto de interesse com maior número de ocorrências em seu redor?", "Qual a distribuição do número de ocorrências pelos diversos pontos de interesse?", "Qual o comportamento temporal de um determinado ponto de interesse?", "Qual a tipologia mais comum em torno de um determinado ponto de interesse" e "Onde acontecem as ocorrências em redor de um ponto de interesse?"

A figura 3.49 apresenta quais as visualizações deste *dashboard* que permitem saber os pontos enumerados acima. Cada número presente na figura corresponde ao número do enumerado. Desta forma é possível fazer um mapeamento entre as visualizações e algumas das questões (enumeradas acima) que estas são capazes de responder.

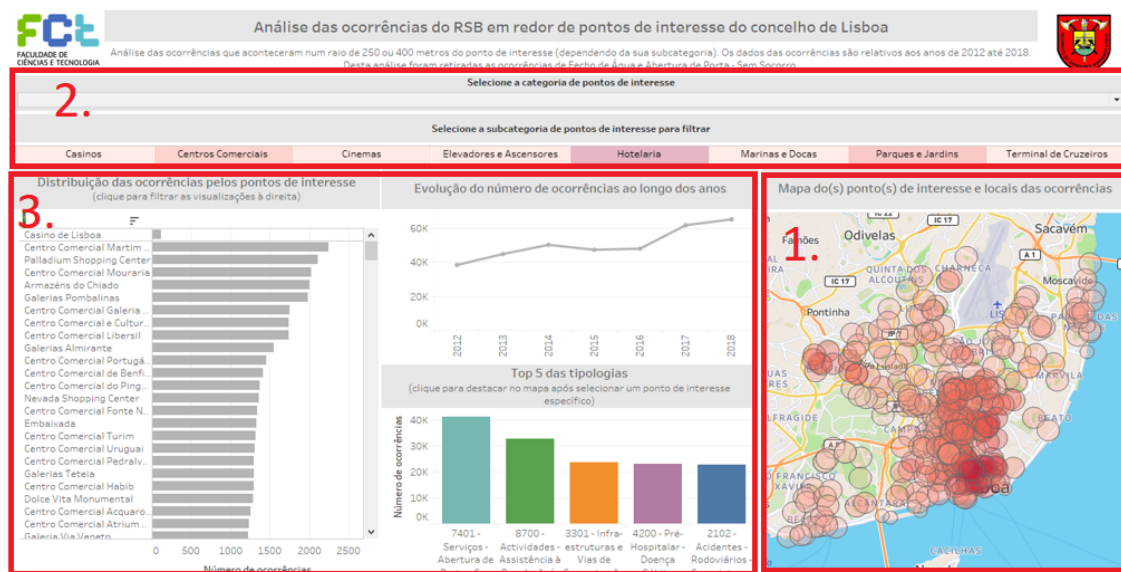


Figura 3.49: Imagem do *dashboard* de análise das ocorrências em redor de pontos de interesse no concelho de Lisboa com indicação das visualizações que respondem às questões.

### 3.5.8 Procura de variáveis explicativas para as ocorrências do RSB - Meteorologia

Os *dashboards* anteriormente apresentados servem para a análise e deteção de padrões nos dados das ocorrências do RSB. No entanto, falta detetar variáveis potencialmente explicativas para os padrões que possam surgir nos dados das ocorrências. Para tal, foi desenvolvido o *dashboard* que procura variáveis explicativas para as ocorrências em relação a dados meteorológicos. Esta ferramenta pretende endereçar a questão relevante "O estado da meteorologia tem impacto no registo de ocorrências no concelho de Lisboa?". Este *dashboard* faz uso dos dados meteorológicos do IPMA anteriormente descritos em 3.3. Para o desenvolvimento desta ferramenta houve a necessidade de dividir os valores possíveis das várias características meteorológicas (humidade, precipitação, temperatura e intensidade do vento) e dar um significado a essa divisão de valores. As tabelas 3.16, 3.17, 3.18 e 3.19 apresentam a divisão feita dos valores para a humidade, precipitação, temperatura e intensidade do vento, respetivamente.

Tabela 3.16: Caracterização dos valores relativos à humidade.

Designação	Intervalo de Valores (%)
Humidade Baixa	< 40
Humidade Normal	[40;70]
Humidade Alta	> 70

A figura 3.50 apresenta uma imagem do *dashboard* desenvolvido que utiliza os dados do IPMA e das ocorrências do RSB e um vídeo demonstrativo das suas funcionalidades está disponível [aqui](#).

Tabela 3.17: Caracterização dos valores relativos à precipitação.

Designação	Intervalo de Valores (mm)
Sem chuva	= 0
Chuva fraca	]0;0.5[
Chuva moderada	[0.5;4.0]
Chuva forte	> 4.0

Tabela 3.18: Caracterização dos valores relativos à temperatura.

Designação	Intervalo de Valores (°C)
Muito frio	(-∞;8[
Frio	[8;14[
Ameno	[14;19[
Agradável	[19;25[
Morno	[25;30[
Quente	[30;36]
Muito quente	> 36

Tabela 3.19: Caracterização dos valores relativos à intensidade do vento.

Designação	Intervalos de Valores (km/h)
Sem vento	= 0
Vento fraco	]0;15[
Vento moderado	[15;36[
Vento forte	[36;56[
Vento muito forte	[56;76[

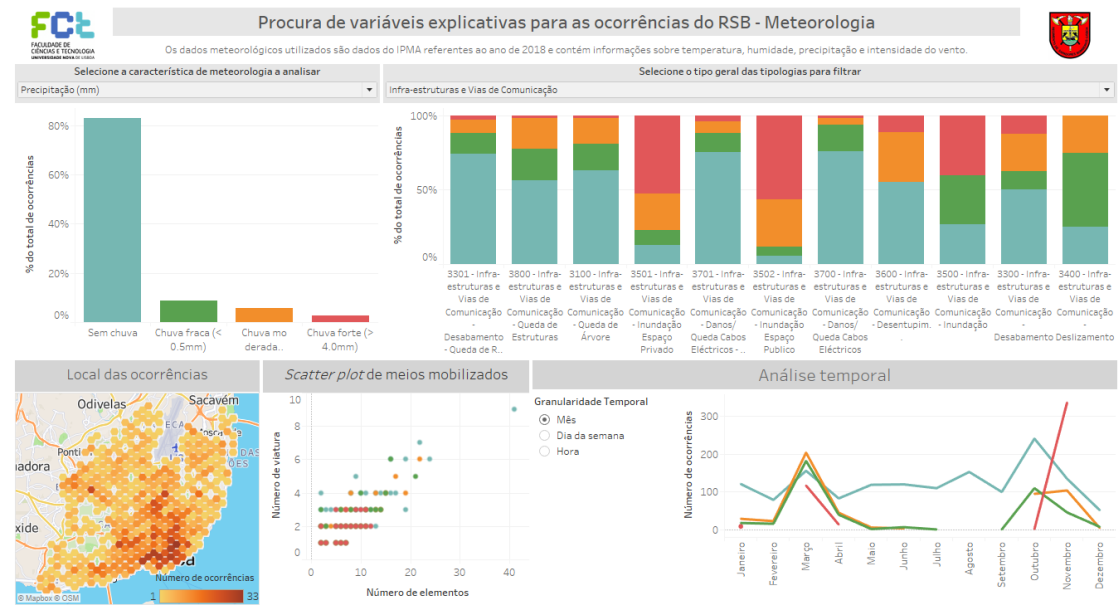


Figura 3.50: Diagrama IVML de tooltips do dashboard de análise das ocorrências em redor de pontos de interesse no concelho de Lisboa.



Para descrever os componentes e as funcionalidades deste *dashboard* estão presentes os diagramas *IVML* de interações e *tooltips* nas figuras 3.51 e 3.52, respectivamente.

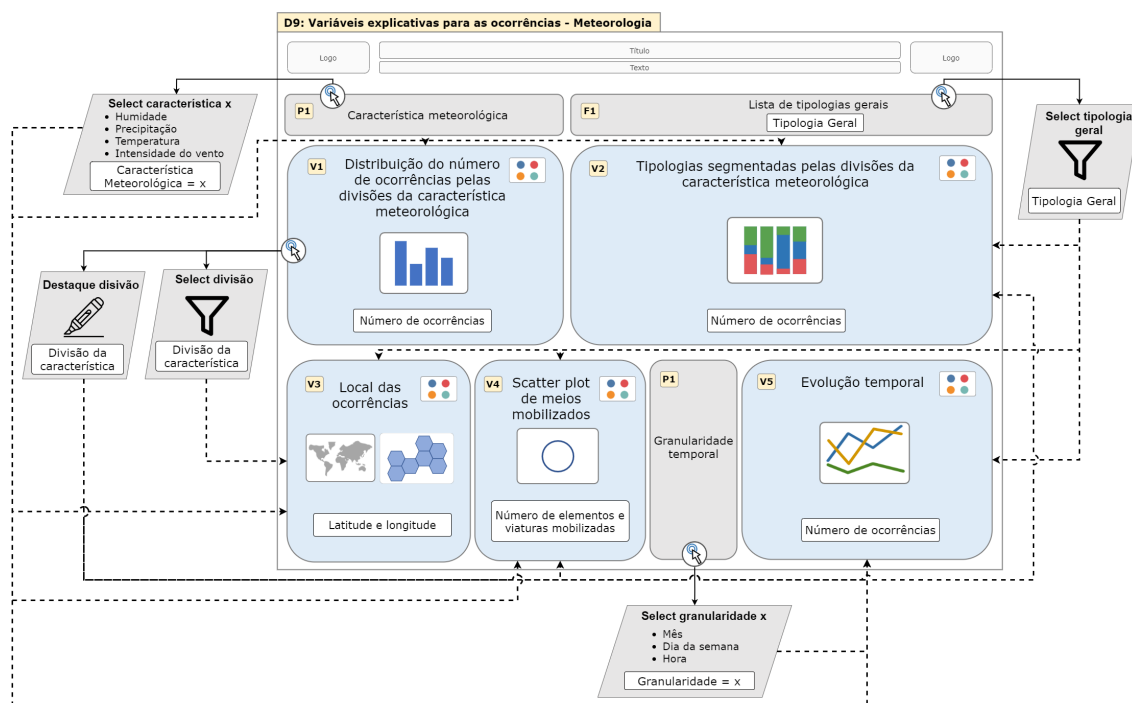


Figura 3.51: Diagrama *IVML* de interações do *dashboard* da procura de variáveis explicativas para as ocorrências do RSB - Meteorologia.

Como é apresentado na figura 3.51, este *dashboard* é composto por dois parâmetros, um filtro e cinco visualizações. O parâmetro P1 apresenta quais são as características meteorológicas disponíveis para análise. O parâmetro P2 apresenta quais são as granularidades temporais disponíveis para análise temporal. Relativamente ao único filtro deste *dashboard*, este é o filtro F1. O filtro apresenta uma lista com as tipologias gerais disponíveis no conjunto de dados das ocorrências. Quantos às visualizações, a visualização V1 apresenta a distribuição do número de ocorrências pelas divisões feitas nas várias características meteorológicas e explicadas acima. Esta visualização é um gráfico de barras e faz uso da cor para mapear para as diferentes divisões consideradas. Isto é, existe uma cor diferente para cada linha das tabelas 3.16, 3.17, 3.18 e 3.19. Este mapeamento de cor tem uma lógica associada, cores como o vermelho e o verde servem para identificar as piores e melhores condições meteorológicas. Estas cores irão ser utilizadas em algumas das outras visualizações e servem como elemento identificador. A visualização V2 apresenta as tipologias divididas pelas divisões consideradas para cada característica do tempo, ou seja, esta visualização permite ver em que condições meteorológicas acontecem as ocorrências de determinadas tipologias. Esta visualização é um gráfico de barras *stacked* a 100%, portanto cada barra que identifica uma tipologia, está dividida em vários segmentos que representam as diferentes divisões meteorológicas consideradas. Esta visualização faz uso da cor para distinguir estes segmentos existentes nas barras e o mapeamento é igual ao

utilizado em V1. A visualização V3 é um mapa de *hexabins* que pretende mostrar o local das ocorrências. Esta visualização faz uso da cor para mapear o número de ocorrências, onde cores mais escuras representam locais com maior concentração de ocorrências. A visualização V4 serve de análise para os meios mobilizados para as ocorrências. É um *scatter plot* do número de elementos e número de viaturas enviadas para o local das ocorrências. Esta visualização faz uso da cor para mapear as diferentes divisões das características meteorológicas. Por fim, a visualização V5 apresenta a evolução temporal das ocorrências. Esta visualização é um gráfico de linhas que apresenta uma linha para cada divisão das características do tempo e utiliza a cor para as identificar.

Relativamente às interações presentes neste *dashboard*, a figura 3.51 mostra que existem duas ações de seleção de parâmetros, duas ações de filtragem e uma de destaque. Uma das ações de escolha de parâmetros é a seleção da característica meteorológica para análise. Esta ação inicia-se quando se seleciona numa das características presentes em P1 e afeta todas as visualizações do *dashboard*. A outra ação de escolha de parâmetros é a escolha da granularidade temporal. Esta ação foi utilizada em vários outros *dashboards* descritos. Inicia-se quando se seleciona a granularidade temporal desejada e apenas afeta a visualização V5. É possível filtrar os dados por várias tipologias gerais. Esta filtragem é possível através da adição ou remoção de tipologias gerais presentes no filtro F1. Esta filtragem irá afetar as visualizações V2 a V5, sendo que estas passarão a apresentar apenas informações relativas às ocorrências das tipologias gerais selecionadas. A última ação de filtragem está ligada à ação de destaque, isto porque, a ação inicia-se com um clique numa das barras presentes na visualização V1 e a ação de filtragem apenas irá afetar a visualização V3, passando esta a apresentar a localização das ocorrências que aconteceram nas condições meteorológicas selecionadas, e a ação de destaque irá afetar a visualização V2, V4 e V5. A ação de destaque irá destacar as divisões da característica meteorológica selecionadas em todas as visualizações que utilizam a cor para as identificar, a visualização V2, V4 e V5.

As *tooltips* presentes neste *dashboard*, como apresentado na figura 3.52, são apenas textuais. A *tooltip* T1 é referente à visualização V1. Esta aparece no ecrã quando se faz *hover* numa das barras e apresenta o valor do número de ocorrências que aconteceram naquela determinada condição meteorológica e a sua percentagem do total de ocorrências. A *tooltip* T2 apresenta dados referentes à visualização V2. As informações aparecem no ecrã quando se faz *hover* do rato num dos segmentos de uma das barras de V2. As informações sobre as ocorrências da tipologia que aconteceram em determinadas condições meteorológicas são o número de ocorrências e a percentagem do total de ocorrências da tipologia. A *tooltip* T3 pertence à visualização V3 e aparece no ecrã quando se faz *hover* num dos hexágonos presentes nessa visualização. As informações apresentadas são o número de ocorrências que estão contidas no hexágono e a percentagem do total de ocorrências que o hexágono contém. A *tooltip* T4 é disponibilizada quando se faz *hover* do rato num dos pontos da visualização V4. São apresentadas as seguintes informações: número de elementos e número de viaturas mobilizadas para o local. Por fim, a última *tooltip* diz



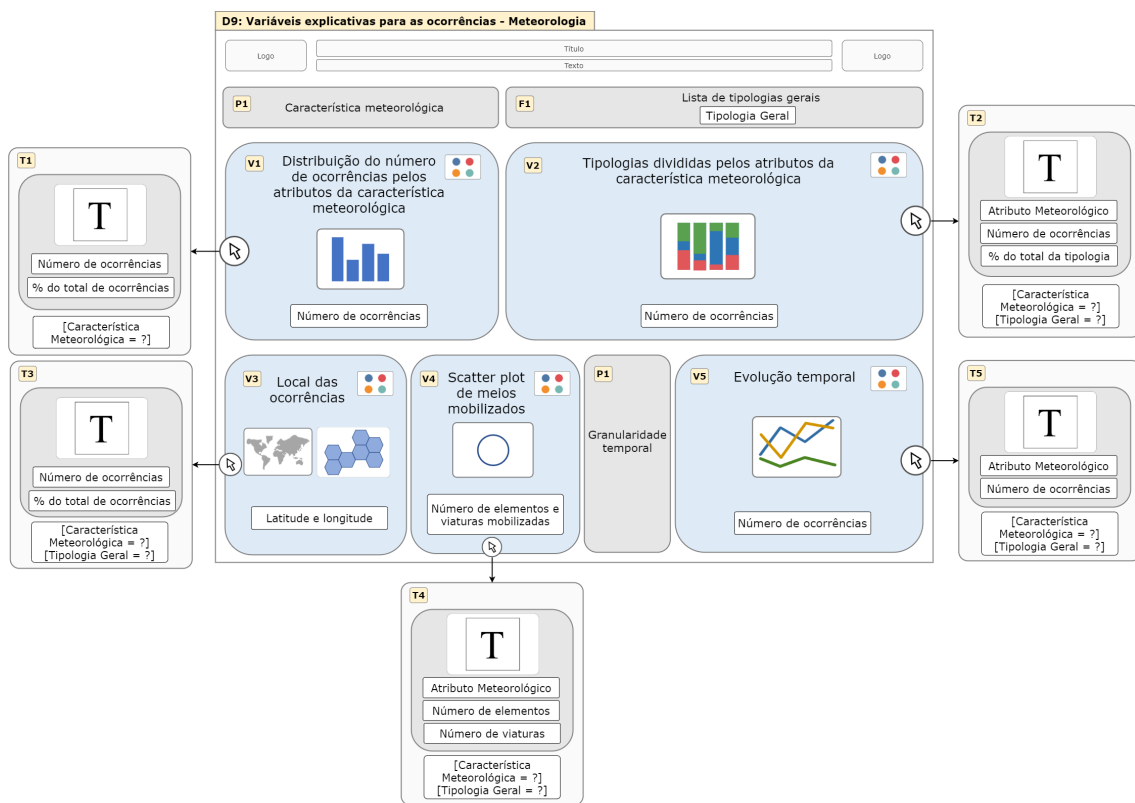


Figura 3.52: Diagrama IVML de tooltips do dashboard da procura de variáveis explicativas para as ocorrências do RSB - Meteorologia.

respeito à visualização V5 e é disponibilizada ao utilizador quando se faz *hover* num dos pontos de uma das linhas presentes nessa visualização. A informação disponibilizada é o número de ocorrências.

Para concluir, com esta ferramenta desenvolvida é possível:

- Observar a distribuição das ocorrências ao longo das condições meteorológicas. Isto permite obter resposta a questões como "Sob que condições meteorológicas existiu mais ocorrências?";
- Analisar o comportamento das tipologias relativamente às características meteorológicas e respetivas divisões. Esta funcionalidade disponibilizada ao utilizador permite obter resposta a perguntas como "Existe alguma tipologia que aconteça mais sob condições meteorológicas mais severas?" e "Quais são as tipologias cujas ocorrências aparentam não estar relacionadas com fenómenos meteorológicos?";
- Observar onde as ocorrências acontecem sobre determinadas condições meteorológicas. Isto permite o utilizador obter respostas a questões como "Quando está chuva forte onde acontecem as ocorrências?" ou "Quando está muito calor onde acontecem as ocorrências?";

- Analisar se existe alguma relação entre o número de elementos e viaturas mobilizadas para o local das ocorrências e o estado do tempo. Perguntas como "Será que quando está chuva forte as ocorrências mobilizam mais meios do que quando está chuva moderada?" poderão ser respondidas através deste *dashboard*;
- Observar o comportamento temporal as ocorrências das várias divisões das características meteorológicas. Isto permite responder a questões como "Qual o valor máximo de ocorrências que aconteceram com chuva forte? E qual o mês, dia da semana ou horas onde isso aconteceu?";
- Com os pontos acima referidos, é possível detetar variáveis potencialmente explicativas para as ocorrências do RSB utilizando diferentes características meteorológicas. Isto possibilita obter resposta às várias questões acima referidas, mas para diferentes características meteorológicas.

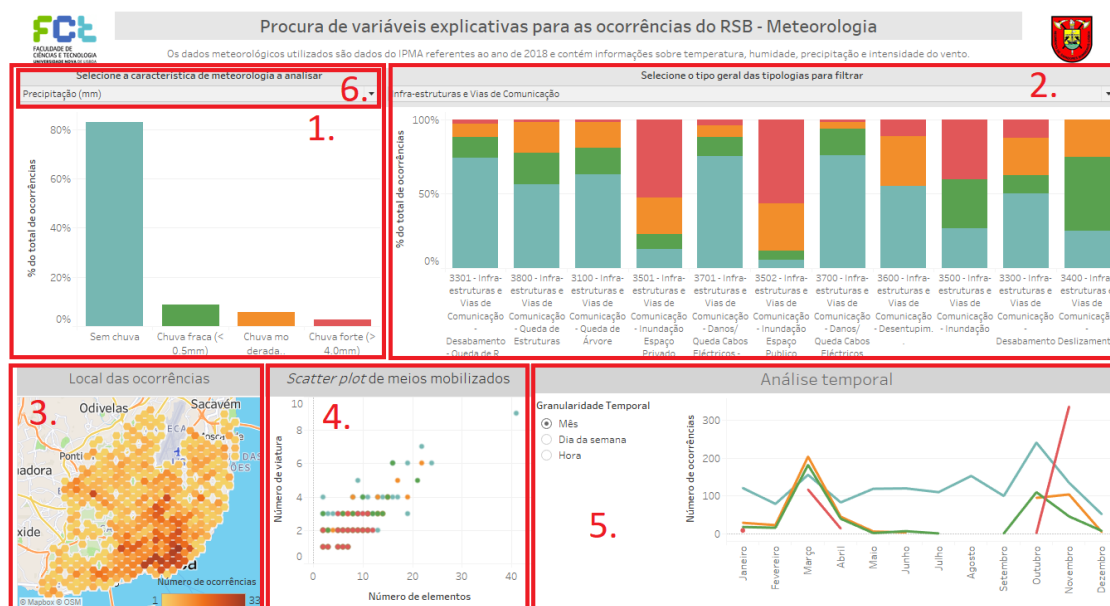


Figura 3.53: Imagem do *dashboard* de procura de variáveis explicativas meteorológicas para as ocorrências do RSB com indicação das visualizações que respondem às questões.

A figura 3.53 apresenta quais as visualizações deste *dashboard* que permitem saber os pontos enumerados acima. Cada número presente na figura corresponde ao número do enumerado. Desta forma é possível fazer um mapeamento entre as visualizações e algumas das questões (enumeradas acima) que estas são capazes de responder.

### 3.5.9 Conclusões

Para obter as respostas às questões levantadas inicialmente na seção 3.1, as questões relevantes, foi necessário desenvolver várias ferramentas visuais que permitem a análise,

exploração e comparação de dados. Com estes desenvolvimentos o Regimento de Sapadores Bombeiros fica munido de ferramentas que ajudarão a explorar, compreender e comparar os dados históricos das ocorrências, descobrir padrões temporais, espaciais e espaço-temporais nos dados das ocorrências, explorar factores externos aos dados das ocorrências, como é o caso do que acontece em redor de pontos de interesse e o que acontece em determinadas condições meteorológicas.

As ferramentas desenvolvidas aparecem na forma de *dashboards* que apresentam um conjunto de visualizações e interações. Cada *dashboard* está responsável por responder a uma das questões relevantes. O processo de desenvolvimento destas ferramentas teve várias iterações, isto é, eram apresentados a professores e alunos de dissertação com o objetivo de retirar comentários e melhorar as ferramentas criadas. Inclusive, os *dashboards* de comparação das ocorrências nas áreas de influência, comparação das ocorrências nas freguesias de Lisboa e evolução espacial das tipologias ao longo do tempo foram apresentados durante uma reunião *online* onde estiveram presentes as duas entidades interessadas, a Câmara Municipal de Lisboa e o Regimento de Sapadores Bombeiros. Após essas reuniões/interações, tanto com os interessados como com os professores e alunos, eram feitas alterações aos *dashboards* de maneira a satisfazer certos requisitos, adicionar funcionalidades, tornar o entendimento e utilização mais fácil ou melhorar a interatividade das ferramentas. Os *dashboards* criados contêm sempre dicas de utilização, por exemplo, sempre que uma visualização serve para iniciar alguma ação, seja de destaque, seja de filtragem, é acrescentado ao título dessa visualização texto que alerta o utilizador para a existência dessa funcionalidade. Um texto de alerta de interações aparece na forma de: "Clique numa tipologia para destacar" ou "Clique num grupo para filtrar todas as visualizações". Quanto às visualizações escolhidas, estas tinham de ser simples de interpretar e acessíveis a todas as pessoas. Isto é, não poderiam ser visualizações muito complexas nem para determinados contextos específicos. As visualizações utilizadas foram escolhidas devido à sua ampla utilização no dia a dia e, conseqüentemente, as pessoas estão familiarizadas com as mesmas e conseguem interpretar corretamente e facilmente. Por exemplo, a utilização de caixas de bigodes para demonstrar a variação dos valores do número de elementos ou viaturas enviadas para o local da ocorrência não seria uma boa escolha de visualização, pois esta utiliza conceitos de estatística que, nem sempre, são compreendidos.

Concluindo, todas as questões levantadas foram resolvidas através do desenvolvimento de *dashboards* interativos e de fácil uso e compreensão que apresentam visualizações com as quais existe muita familiarização. A escolha das visualizações utilizadas foi pensada, sendo apenas escolhidas visualizações de fácil compreensão e perceção. As ferramentas criadas contêm dicas sobre as funcionalidades existentes para facilitar a vida do utilizador. Com a criação destas ferramentas o Regimento de Sapadores Bombeiros pode retirar várias informações delas, sendo que pode fazer a deteção de padrões temporais, espaciais e espaço-temporais.

### 3.6 Algumas Análises com as Ferramentas Desenvolvidas

Ao longo do desenvolvimento desta dissertação, foi-se analisando os dados das ocorrências com as ferramentas descritas na seção 3.5. Nesta seção serão abordadas algumas das análises com mais interesse e relevância e que exemplificam o uso e descobertas que podem ser feitas com as ferramentas desenvolvidas.

O exemplo de análise presente na figura 3.54 demonstra uma análise das ocorrências de tipologia "Pré-Hospitalar" utilizando o *dashboard* da evolução espacial das tipologias ao longo dos anos. Com esta figura é possível observar que durante os anos de 2012, 2013, 2014 e 2015 a maioria das ocorrências deste tipo estavam concentradas na zona de Benfica. No ano de 2016 é possível visualizar que houve uma mobilização da concentração das ocorrências da zona de Benfica para a zona do Parque das Nações/Olivais e que nos anos de 2017 e 2018 a concentração das ocorrências "Pré-Hospitalares" moveu-se novamente para outra região do concelho, para a baixa da cidade de Lisboa. Esta análise serve para consolidar que esta ferramenta desenvolvida é capaz de apresentar a existência de *hotspots* entre os vários anos existentes nos dados e que é capaz de expressar a movimentação desses *hotspots* ao longo do tempo.

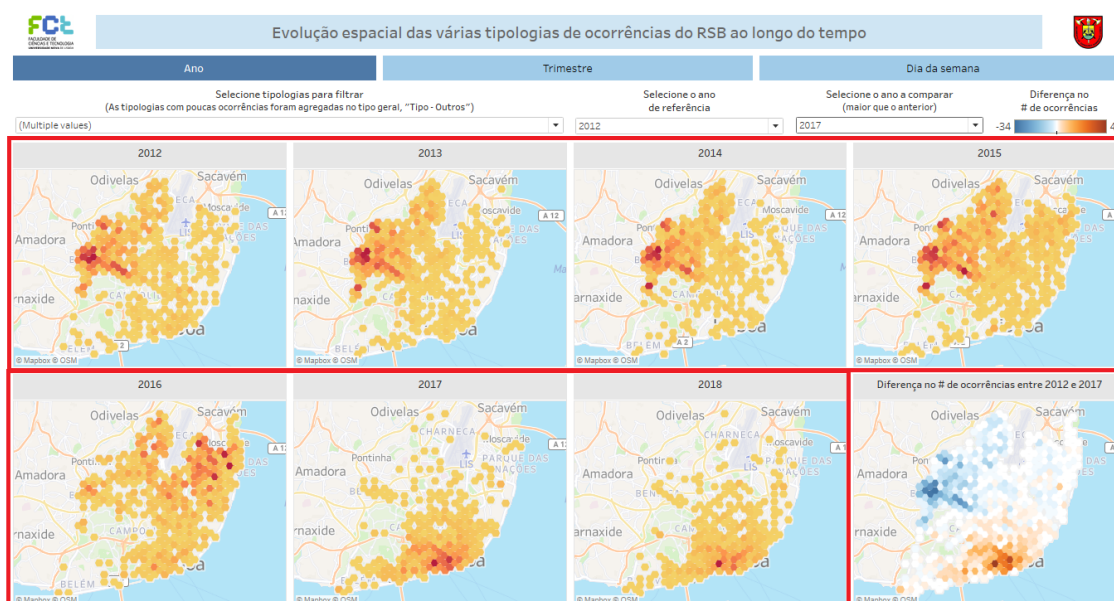


Figura 3.54: Exemplo de análise do *dashboard* da evolução espacial das tipologias ao longo dos anos.

Outra análise disponibilizada por este *dashboard* é uma análise ao nível dos trimestres. A figura 3.55 apresenta uma imagem do *dashboard* da evolução espacial das tipologias "Incêndios" ao longo dos trimestres. Esta análise é curiosa. Tipicamente quando se pensa em incêndios associa-se a tempo quente e meses de verão. No entanto, os resultados obtidos na figura 3.55 demonstram que as ocorrências de incêndios acontecem durante todo o ano. O primeiro trimestre (Janeiro, Fevereiro e Março) tem um comportamento

semelhante ao quarto trimestre (Outubro, Novembro e Dezembro), as ocorrências de incêndios acontecem de forma dispersa pelo concelho de Lisboa. O segundo trimestre (Abril, Maio e Junho) já evidencia a existência de um *hotspot* na zona entre o Beato e Marvila, sendo que existe ainda alguma dispersão das ocorrências. O terceiro trimestre do ano (Julho, Agosto e Setembro) apresenta claramente um *hotspot* também na zona de Beato e Marvila. A dispersão observada no primeiro e último trimestre, que são os mais frios, poderá sugerir incêndios de carácter doméstico, por exemplo, devido ao uso inadequado de aquecedores. Novamente, esta análise consolida que a utilização destas visualizações é eficaz na deteção de *hotspots* e verificar a sua movimentação ao longo de várias granularidades temporais.

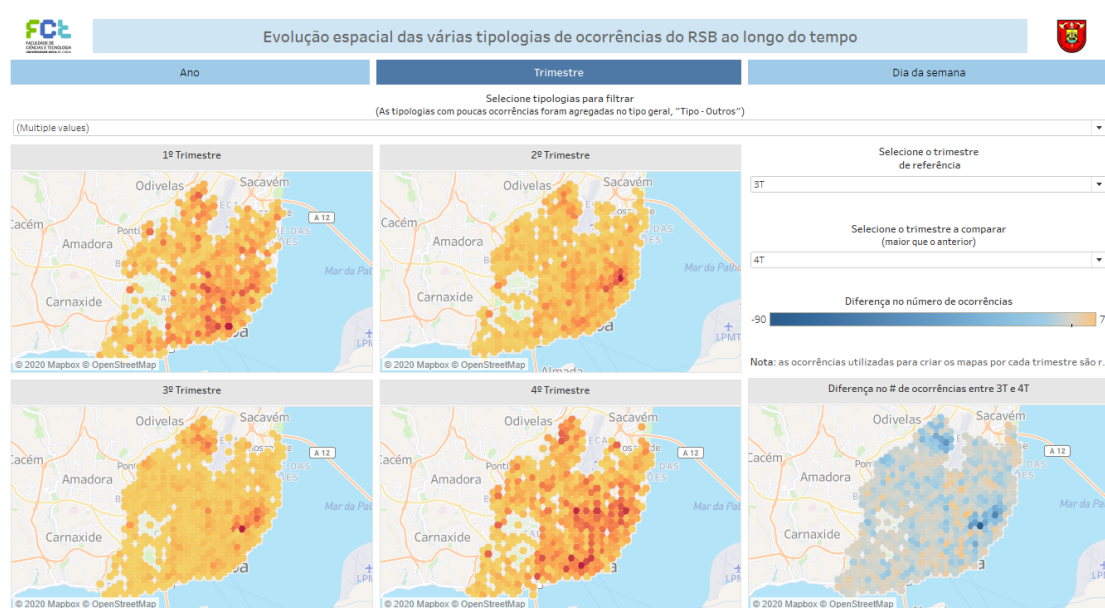


Figura 3.55: Exemplo de análise do *dashboard* da evolução espacial das tipologias ao longo dos trimestres.

A análise anteriormente feita para as ocorrências de tipologia "Pré-Hospitalar" causou algum interesse e pode ser explorada através de outras ferramentas desenvolvidas, nomeadamente, o *dashboard* de comparação das ocorrências nas áreas de influência dos quartéis do RSB. A figura 3.56 ilustra a comparação das áreas de influências de Benfca, Quartel do Comando e Praça da Alegria utilizando esse *dashboard*. A escolha destas áreas de influência recaiu na análise anteriormente feita que demonstrou a mobilização das ocorrências "Pré-Hospitalares" da zona de Benfca para a baixa da cidade de Lisboa. Como é possível observar na figura 3.56, no gráfico de linhas responsável pela análise temporal das ocorrências das áreas de influência, a área de influência de Benfca a partir do ano de 2015 apresentou um grande decréscimo da média diária de ocorrências "Pré-Hospitalares". Em 2015, em média existiam cerca de três ocorrências deste tipo por dia na área de influência de Benfca e em 2016 a média nem chega ao valor de 0.5. Em contraste, as áreas de influências Quartel do Comando e Praça da Alegria aumentaram a média de

ocorrências "Pré-Hospitalares" diárias a partir do ano de 2016. A ideia de que as ocorrências que deixaram de acontecer na zona de Benfica e passaram para a baixa da cidade de Lisboa é suportada tanto por estes resultados como pelos apresentados anteriormente no *dashboard* da evolução espacial das tipologias ao longo do tempo. Esta análise exemplifica uma análise de comparação entre várias áreas de influência, nomeadamente, em termos de características temporais.

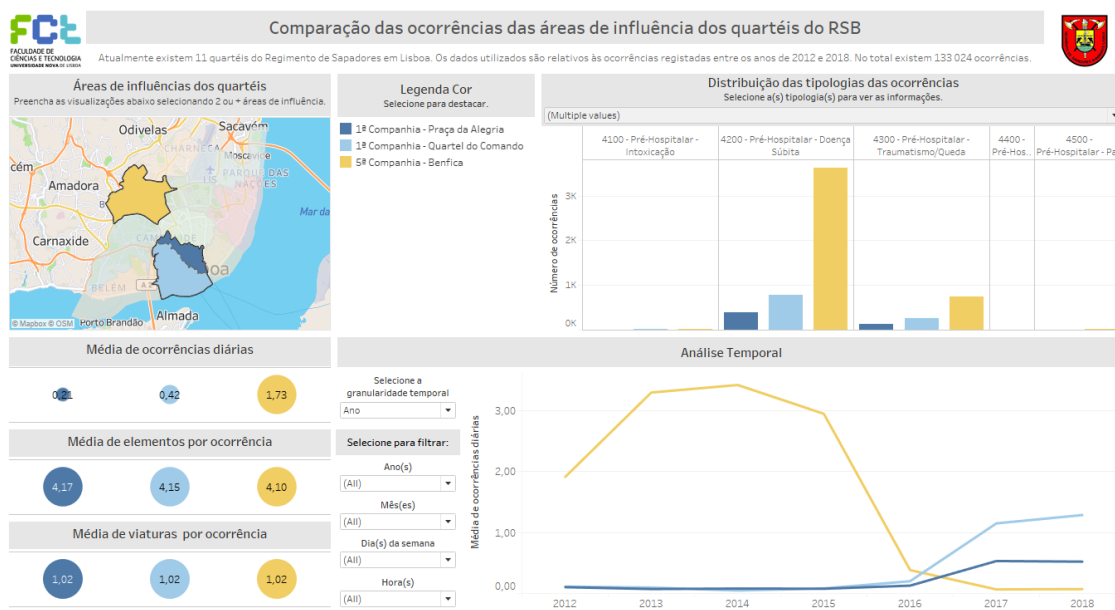


Figura 3.56: Exemplo de análise do *dashboard* de comparação das ocorrências nas áreas de influência dos quartéis.

A análise feita anteriormente para as tipologias do tipo de "Incêndios", relativamente à sua evolução ao longo dos trimestres, levantou alguma curiosidade, por isso recorreu-se ao *dashboard* de análise do comportamento temporal das tipologias ao longo do tempo. A figura 3.57 apresenta esse *dashboard* após ter sido feito uma ação de destaque do grupo 1. Esta figura permite evidenciar o comportamento temporal, ao longo das estações do ano, do grupo 1 e observar onde as ocorrências desse grupo estão espacialmente. É visível que as tipologias presentes neste grupo apresentam uma sazonalidade, particularmente na época de Verão onde atinge o pico do número de ocorrências. Olhando para o mapa de *hexabins* disponibilizado por este *dashboard* podemos observar que existe uma grande concentração das ocorrências referentes às tipologias presentes neste grupo na região de Monsanto. No entanto, existe também alguma concentração na zona entre o Beato e Marvila. A lista de tipologias que partilham este comportamento de sazonalidade no Verão está presente na tabela denominada "Tipologias". Como seria de esperar, as tipologias que apresentam este comportamento temporal são, em grande parte, tipologias associadas a incêndios. Esta análise interliga-se com a anteriormente efetuada para as tipologias relacionadas com "Incêndios" no sentido em que um dos *hotspots* deste grupo temporal é na mesma região que o *hotspot* detatado atrás, na zona entre o Beato e Marvila. Com



esta investigação é exemplificada uma forma de analisar os resultados que este *dashboard* apresenta.

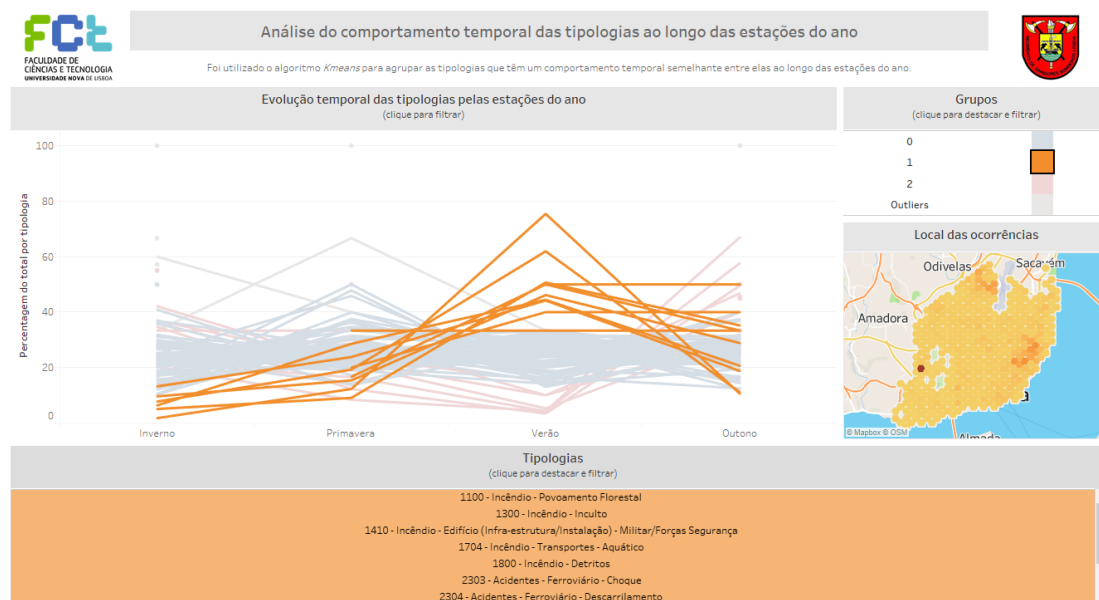


Figura 3.57: Exemplo de análise do *dashboard* de análise do comportamento temporal das tipologias do grupo 1 ao longo das estações do ano.

Ainda a analisar os resultados apresentados no *dashboard* de análise do comportamento temporal de tipologias ao longo das estações do ano, a figura 3.58 demonstra o que é apresentado neste *dashboard* após a seleção do grupo 2 para destaque e análise. A figura 3.58 demonstra que o comportamento temporal deste grupo também apresenta sazonalidade. As ocorrências das tipologias contidas neste grupo são mais comuns na estação de Inverno, onde existe um pico no número de ocorrências, e são menos comuns na época de Verão, onde existe o ponto mais baixo. As ocorrências contidas neste grupo estão mais concentradas na zona da baixa da cidade de Lisboa. Isto pode ser observado no mapa de *hexabins*. Como esperado, é observável, através da lista presente na tabela "Tipologias", que as tipologias que têm esta sazonalidade no Inverno são tipologias de "Infra-estruturas e Vias de Comunicação", ou seja, desabamentos, deslizamentos e inundações. Estas tipologias poderão estar relacionadas com fenómenos de mau tempo. No entanto, o que não seria de esperar era que tipologias de incêndios, nomeadamente, agrícolas partilhassem este comportamento temporal com tipologias normalmente associadas a mau tempo. Este *dashboard* permite fazer este tipo de análises interessantes e demonstra que existem alguns resultados que não são expectáveis inicialmente.

A figura 3.59 apresenta o *dashboard* de análise do comportamento espacial das ocorrências do RSB após a seleção de dois grupos para destaque e filtragem na lista que apresenta os grupos resultantes do algoritmo de agrupamento utilizando apenas as ocorrências consideradas de emergência. Os grupos espaciais selecionados são os grupos 0 e 11. Ao analisar a figura 3.59 é visível que o grupo 0 situa-se na região de Campolide e o grupo

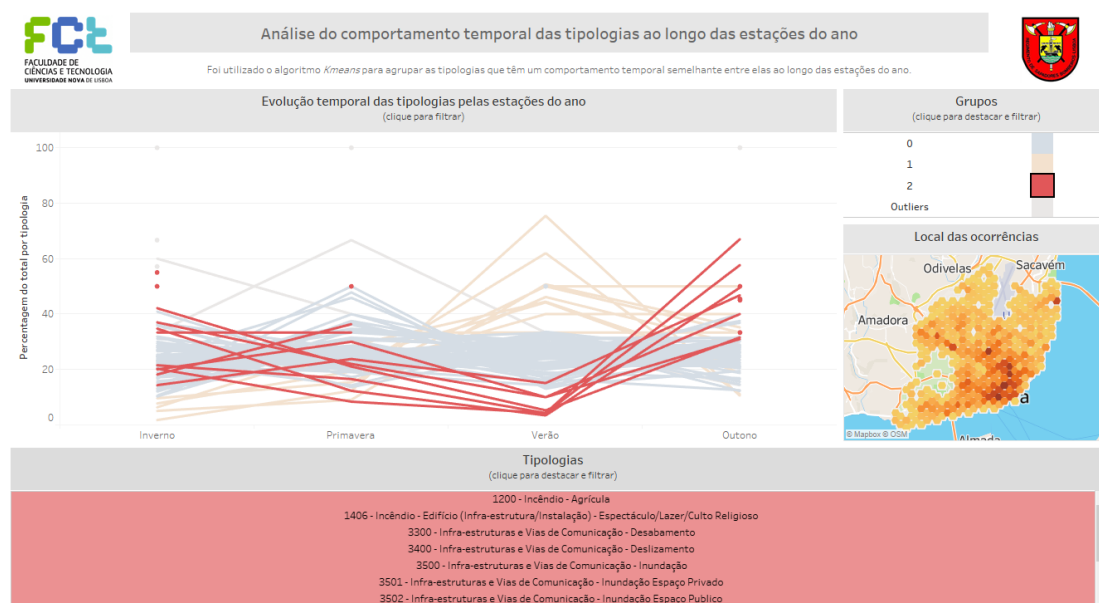


Figura 3.58: Exemplo de análise do *dashboard* de análise do comportamento temporal das tipologias do grupo 2 ao longo das estações do ano.

11 situa-se na região da Charneca. Também é visível, através da leitura do gráfico de barras, que existe uma diferença entre a tipologia mais comum nestes grupos. O grupo 0 apresenta a tipologia de acidentes rodoviários como a mais comum na sua área geográfica, enquanto que, o grupo 11 apresenta a tipologia de incêndios como a mais comum na sua área geográfica. Esta análise permite evidenciar que estes grupos para além de serem diferentes em termos espaciais, são também diferentes em termos das características das suas ocorrências, nomeadamente, as tipologias das ocorrências. Outra análise possível com este *dashboard* é a análise temporal. É visível pelo gráfico de linhas apresentado que o grupo 11, que continha os incêndios como tipologia mais comum, apresenta sazonalidade nos meses de maior calor. Relativamente ao comportamento temporal das ocorrências do grupo 0, este apresenta um ligeiro aumento do número de ocorrências durante os meses de pior tempo. Esta análise pode ser efetuada em diversas granularidades temporais. A figura 3.60 apresenta a análise temporal das ocorrências do grupo 0 e 11 ao longo dos dias da semana. Com esta figura é visível que existe um pico no número de ocorrências do grupo 0 durante a sexta-feira. Visto que as principais ocorrências são do tipo de acidentes rodoviários e que existe um pico durante a sexta-feira, isto poderá levar a pensar que as pessoas estão com pressa de chegar a casa para o fim de semana e aconteçam acidentes rodoviários devido a essas pressas. Relativamente ao grupo 11, este apresenta um pico no número de ocorrências às terças-feiras. Indo ainda mais longe na análise temporal, a figura 3.61 apresenta esta análise ao nível horário. O grupo 0 apresenta um pico no número de ocorrências ao final da tarde, às 18 horas. Isto poderá ainda suportar a ideia de que os condutores têm urgência em chegar a casa e acontecerem acidentes. Quanto ao grupo 11, o comportamento temporal observado é que a partir da uma da manhã a tendência



### 3.6. ALGUMAS ANÁLISES COM AS FERRAMENTAS DESENVOLVIDAS

do número de ocorrências é para decrescer e a partir das cinco da manhã existe uma tendência de crescimento do número de ocorrências. Estas investigações são exemplos de possíveis análises que são permitidas por este *dashboard* de análise do comportamento espacial das ocorrências do RSB.

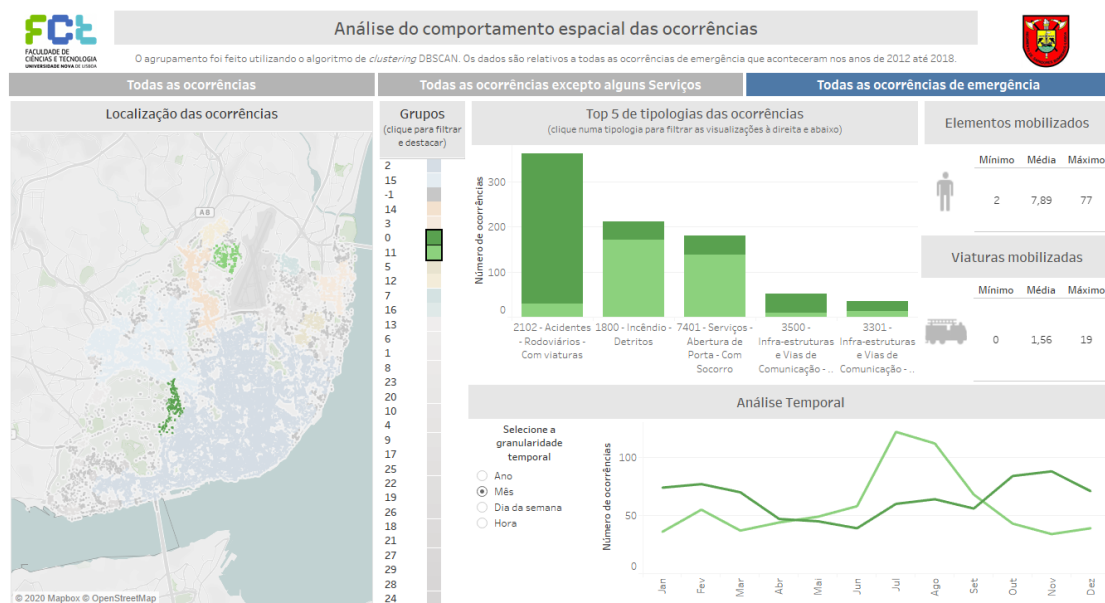


Figura 3.59: Exemplo de análise do *dashboard* de análise do comportamento dos grupos 0 e 11 resultantes do agrupamento espacial das ocorrências de emergência.

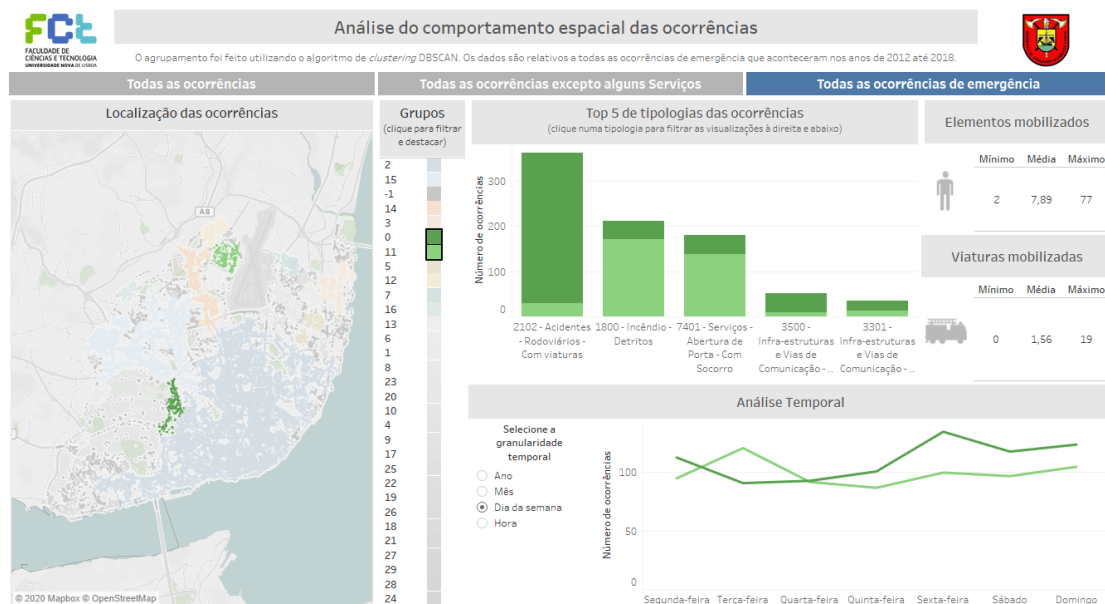


Figura 3.60: Exemplo de análise do *dashboard* de análise do comportamento dos grupos 0 e 11 resultantes do agrupamento espacial das ocorrências de emergência ao longo dos dias da semana.

A figura 3.62 apresenta o *dashboard* de análise das ocorrências que aconteceram em

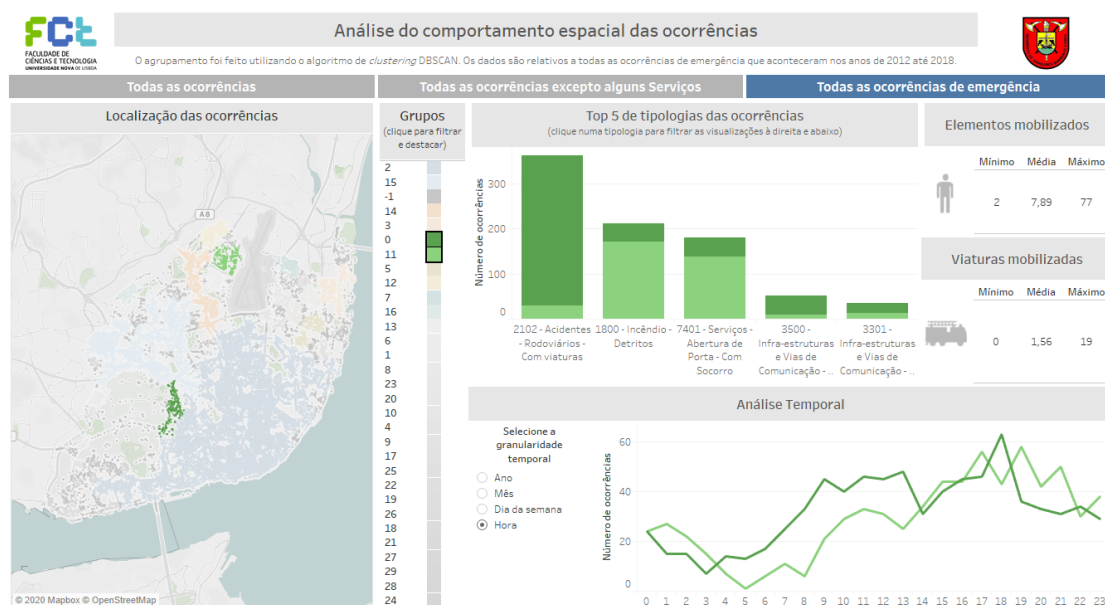


Figura 3.61: Exemplo de análise do *dashboard* de análise do comportamento dos grupos 0 e 11 resultantes do agrupamento espacial das ocorrências de emergência ao longo das horas.

redor de pontos de interesse do concelho de Lisboa. A figura apresenta o *dashboard* após a seleção da categoria de interesse dos pontos de interesse de "Transportes" e após a seleção da subcategoria "Estações Ferroviárias". Assim este *dashboard* apresenta apenas dados de ocorrências que aconteceram em redor de estações ferroviárias. Pelo gráfico de barras é possível visualizar que a estação ferroviária do Rossio é a que apresenta mais ocorrências em seu redor. Este facto também é apoiado pela observação do mapa presente no *dashboard*. É possível observar também que no ano de 2017, relativamente ao primeiro ano de dados (2012), houve uma duplicação do número de ocorrências em redor destes pontos de interesses relativos às "Estações Ferroviárias". Querendo-se analisar o ponto de interesse que contém maior número de ocorrências, a estação do Rossio, a figura 3.63 apresenta o resultado do *dashboard* após a seleção desse ponto de interesse em particular. Com esta figura é possível observar a sua evolução ao longo dos anos, sendo que em 2017 o número de ocorrências registadas triplicou em relação ao primeiro ano de dados. Também é possível ver que a maioria das ocorrências que acontecem em redor da estação ferroviária do Rossio é de assistência e apoio social à população. Por fim, é possível visualizar no mapa onde as ocorrências aconteceram geograficamente e consegue-se entender que existe um aglomerado de ocorrências de assistência à população na zona direita do *buffer*. Este aglomerado de ocorrências poderá ser um potencial alvo de análise. Estas investigações são exemplos de possíveis análises que são permitidas por este *dashboard* de análise das ocorrências em redor dos pontos de interesse do concelho de Lisboa.

### 3.6. ALGUMAS ANÁLISES COM AS FERRAMENTAS DESENVOLVIDAS

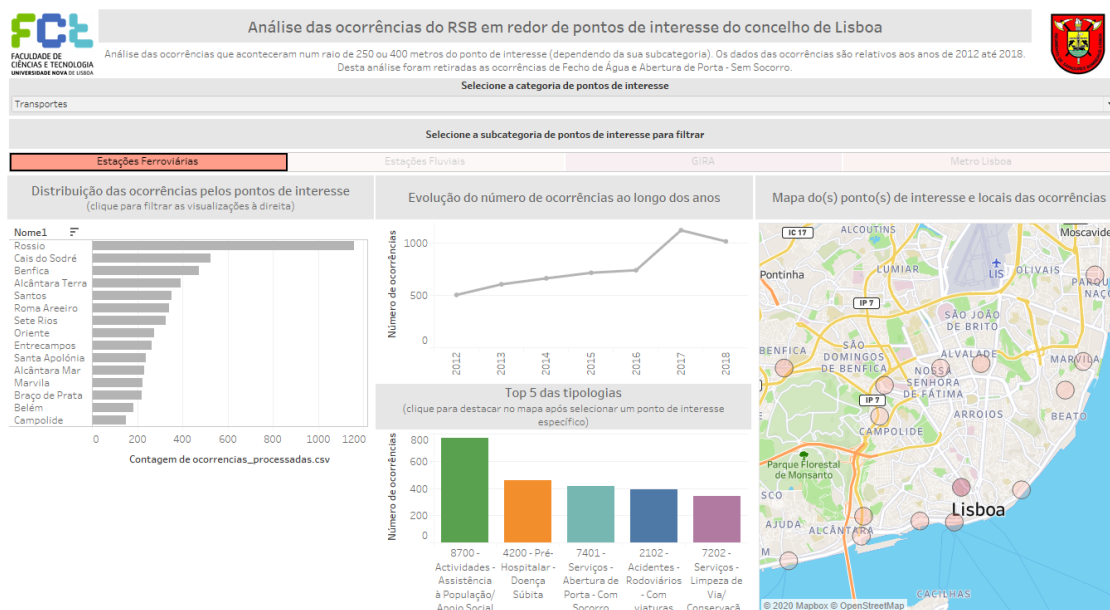


Figura 3.62: Exemplo de análise do *dashboard* de análise das ocorrências em redor de pontos de interesse (transportes).

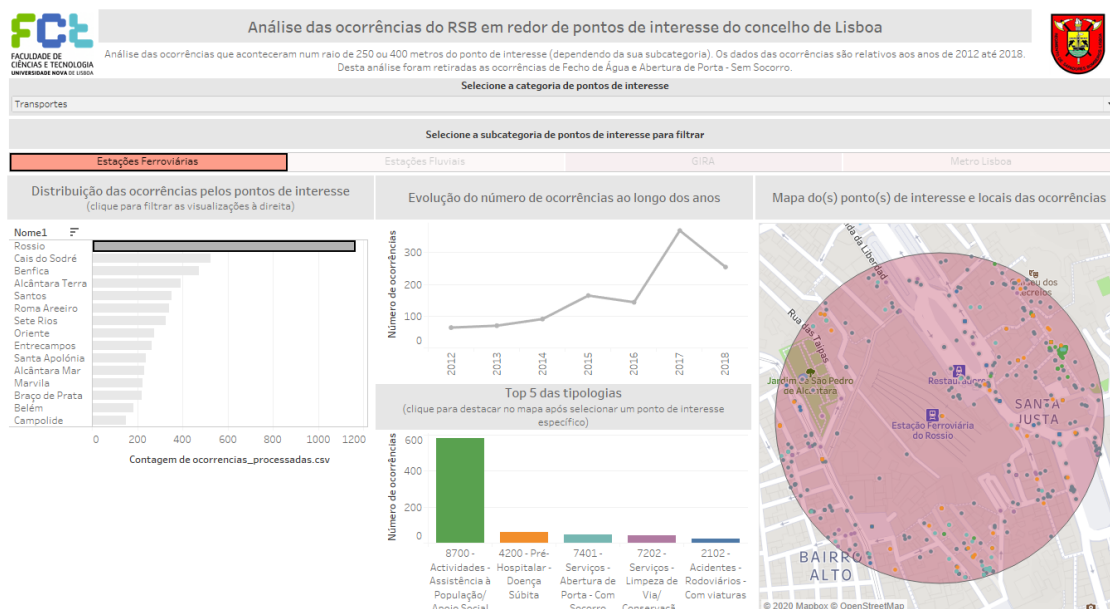


Figura 3.63: Exemplo de análise do *dashboard* de análise das ocorrências em redor de um ponto de interesse.

A figura 3.64 apresenta o *dashboard* da procura de variáveis explicativas para as ocorrências do RSB - Meteorologia. Esta figura apresenta este *dashboard* após a seleção da "Precipitação" como a característica meteorológica de interesse de análise e "Infra-estruturas e Vias de Comunicação" como a tipologia geral de interesse. É possível observar pelo gráfico de barras *stacked* que existem certas tipologias que contêm uma percentagem elevada de ocorrências que aconteceram sob chuva moderada e chuva forte. Essas tipologias são referentes a inundações, desabamentos e desentupimento/tamponamento. As restantes tipologias as ocorrências acontecem quando não está a chover ou quando está chuva fraca. Com esta observação é possível investigar mais sobre as ocorrências que acontecem com chuva moderada e chuva forte. Para tal é necessário selecionar as barras correspondentes a esses valores. A figura 3.65 demonstra o resultado do *dashboard* após essa seleção. Com esta figura é mais perceptível quais as tipologias que têm ocorrências que aconteceram com estas condições meteorológicas. É também possível observar, através do mapa de *hexabins* presente neste *dashboard*, que estas condições meteorológicas afetam mais a zona da baixa e centro de Lisboa. Zonas mais recentes de Lisboa, como por exemplo, o Parque das Nações não sofre tanto com condições meteorológicas mais severas em termos de precipitação. Outra conclusão que podemos retirar desta análise é que as ocorrências que aconteceram com chuva moderada foram as que mais meios humanos e veículos mobilizaram para os locais das ocorrências. Isto é visível no *scatter plot* presente no *dashboard*, onde as ocorrências sob chuva forte mobilizaram, no máximo, cerca de 12 elementos e 3 viaturas e as ocorrências sob chuva moderada mobilizaram, no máximo, cerca de 22 elementos e 6 viaturas. Por fim, é possível fazer a análise temporal destas ocorrências. Sem surpresas estas acontecem nos meses de Outono e Inverno, sendo que as ocorrências que aconteceram com chuva moderada têm um pico no mês de Março e as ocorrências com chuva forte têm um pico acentuado no mês de Novembro. Este é um exemplo de uma análise que pode ser efetuada com o *dashboard* de procura de variáveis explicativas para as ocorrências considerando a meteorologia. Esta análise poderia também ser efetuada para as diferentes características meteorológicas disponíveis.

Estas análises aqui apresentadas são meramente exemplificativas do que se pode descobrir com as várias ferramentas desenvolvidas nesta dissertação, sendo que existem ainda mais análises possíveis que não foram aqui abordadas.

Para concluir, estes exemplos apoiam que algumas das visualizações utilizadas nos *dashboards* servem para responder às diversas questões relevantes. No entanto, é necessário submeter todas estas ferramentas para avaliação com utilizadores.

### 3.6. ALGUMAS ANÁLISES COM AS FERRAMENTAS DESENVOLVIDAS

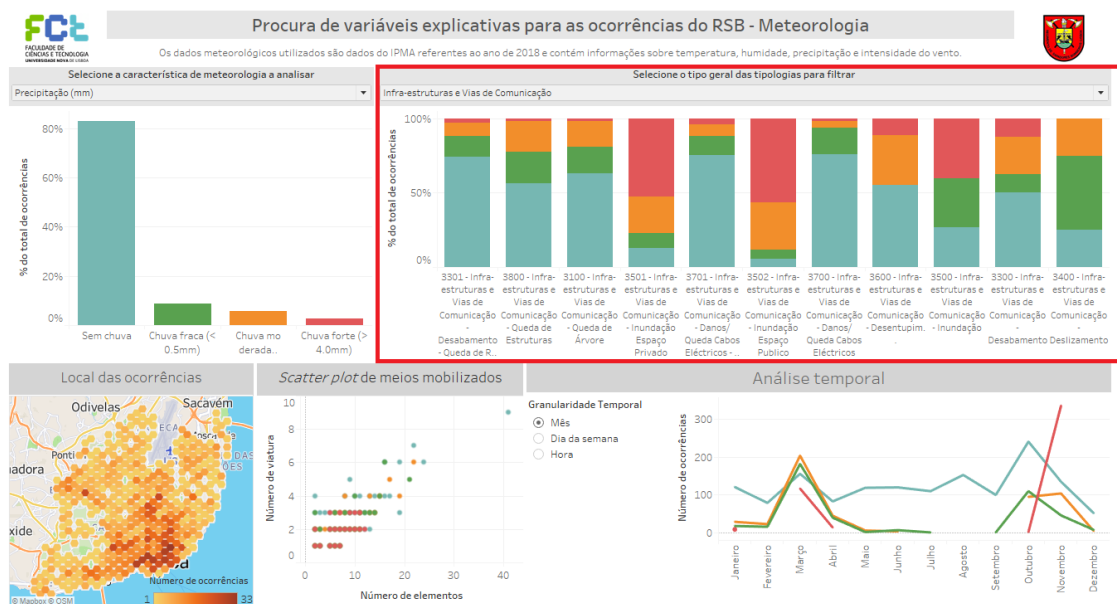


Figura 3.64: Exemplo de análise do *dashboard* de procura de variáveis explicativas para as ocorrências do RSB - Meteorologia.

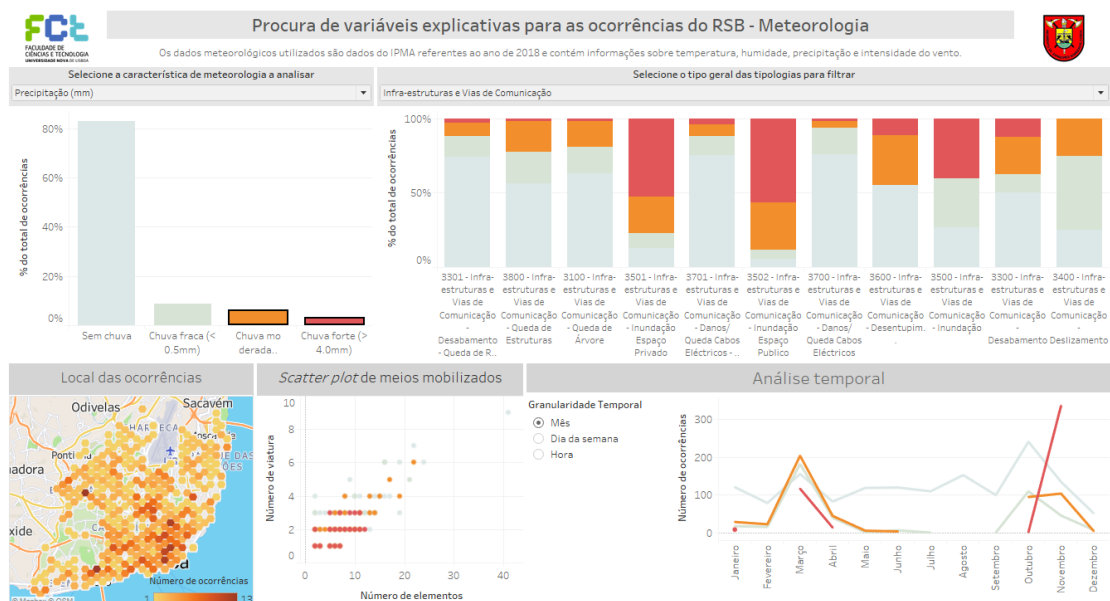


Figura 3.65: Exemplo de análise do *dashboard* de procura de variáveis explicativas para as ocorrências do RSB - Meteorologia - Chuva moderada e chuva forte.



## AVALIAÇÃO DO TRABALHO REALIZADO

Neste capítulo será apresentado o processo de avaliação experimental de todo o trabalho que foi desenvolvido e descrito no capítulo 3.

Este capítulo contém, na seção 4.1, a descrição de todo o processo de avaliação experimental, nomeadamente, qual o processo inicialmente pensado e qual foi o realizado, quais as ferramentas utilizadas para recolher as métricas de utilização. A taxa de sucesso e os tempos de resposta foram utilizados para aferir a qualidade das ferramentas desenvolvidas. A seção 4.2 contém uma descrição dos participantes que realizaram esta experiência. Por fim, na seção 4.3 serão apresentados os resultados obtidos neste processo de avaliação experimental das ferramentas desenvolvidas nesta dissertação.

### 4.1 Avaliação Experimental

Inicialmente, a avaliação do trabalho desenvolvido, nomeadamente, a avaliação das ferramentas criadas durante a dissertação seria feita por utilizadores especializados na temática, ou seja, pessoal do RSB. Para tal, as ferramentas desenvolvidas iam sendo disponibilizadas no *Tableau Online*, onde tinham sido criadas credenciais para os utilizadores acederem às mesmas. Desta forma os utilizadores especializados poderiam interagir com os *dashboards* e explorar as suas funcionalidades. Num cenário ideal, seriam concretizadas reuniões para discutir as ferramentas criadas, receber *feedback* sobre as mesmas ou até apresentadas visualizações alternativas às existentes. Em Setembro foram disponibilizados, no *Tableau Online*, três dos *dashboards* desenvolvidos, nomeadamente, os *dashboards* de comparação das ocorrências nas áreas de influência dos quartéis, de comparação das ocorrências nas freguesias de Lisboa e da evolução espacial das tipologias ao longo do tempo. Estes *dashboards* foram disponibilizados para serem experimentados e para serem discutidos. No entanto, o *feedback* tardou a chegar e quando chegou não era suficiente.



Toda a avaliação das ferramentas dependia da aprovação dos utilizadores especializados na temática, visto estas ferramentas serem desenvolvidas exclusivamente para o caso em questão, as ocorrências do *RSB*, e não terem aplicação a mais nenhum outro cenário.

Devido à situação excecional vivida, a pandemia COVID-19, não foi possível concretizar este plano de avaliação das várias ferramentas desenvolvidas com utilizadores especializados na temática. Houve um esforço tanto da minha parte como da parte dos professores envolvidos neste trabalho em obter respostas e em agilizar o processo de *feedback* das ferramentas que já tinham sido submetidas para avaliação. No entanto, mesmo com este esforço não foi possível obter avaliação dos utilizadores especializados na temática. O *RSB*, durante estes tempos difíceis e atípicos, não conseguiu disponibilizar nem tempo nem recursos humanos para o desenvolvimento desta tarefa de avaliação.

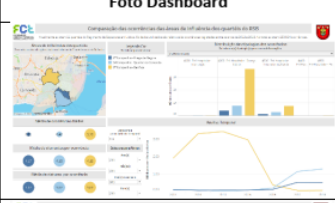


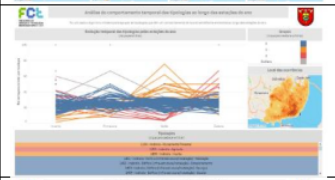
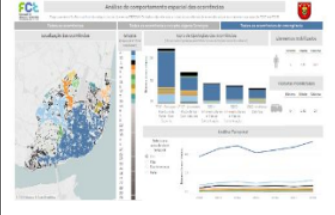
Não sendo possível fazer a avaliação prevista das ferramentas desenvolvidas, houve a necessidade de recorrer a um plano B. Este plano B consiste em fazer a avaliação das visualizações/ferramentas com utilizadores que não são entendidos na temática, isto é, não são utilizadores especializados no tema das ocorrências do *RSB*. Esta avaliação pretende avaliar a taxa de sucesso de cada visualização/ferramenta através de perguntas. Para cada pergunta será obtido o tempo que o utilizador demorou a responder à mesma utilizando a visualização/ferramenta disponibilizada. Também será alvo de análise a taxa de sucesso do utilizador, isto é, se conseguiu responder à pergunta corretamente ou não. Relativamente ao tempo de resposta que será recolhido, este servirá apenas para efeitos de documentação, pois não existem outras ferramentas concorrentes cujos resultados de tempo de resposta do utilizador possam ser utilizados para comparação.

Para executar este método de avaliação é necessário definir quais as visualizações/ferramentas que irão ser submetidas a avaliação e quais as questões que serão consideradas para as avaliar. Sendo este um plano B, houve necessidade de fazer um levantamento das perguntas. Este levantamento foi feito, em parte, com algumas informações das reuniões que inicialmente existiram com o *RSB*, nas quais a entidade expunha os seus interesses, objetivos e questões que gostariam de ver respondidas, mas a maioria das questões foram criadas por mim e sob a direção dos orientadores que participaram nesta dissertação. As tabelas 4.1 e 4.2 apresentam as diversas perguntas que serão apresentadas aos utilizadores num questionário. Estas tabelas apresentam o nome da ferramenta (*dashboard*), quais as visualizações do *dashboard* que serão avaliadas, o identificador da pergunta, a categoria da pergunta e, por fim, qual a questão que será colocada ao utilizador. Como as tabelas 4.1 e 4.2 demonstram, a maior parte das perguntas são referentes ao objetivo de deteção de padrões, da pergunta P01 a P15, e as restantes serão sobre a procura de variáveis explicativas, da tarefa P16 à P23. Cada um dos *dashboards* desenvolvidos e apresentados na seção 3.5 será alvo de avaliação em termos de recolha de métricas como o tempo de resposta e taxa de sucesso da visualização. Alguns desses *dashboards* apresentam os resultados em diferentes granularidades temporais ou em diferentes agrupamentos, mas como são sempre utilizadas as mesmas visualizações para apresentar os resultados não existe a necessidade de avaliar todos esses *dashboards*. É suficiente avaliar um deles. As





questões foram escolhidas de modo a abranger quase todas as visualizações presentes nos *dashboards*, sendo que, uma resposta a uma questão pode ser obtida numa visualização ou num conjunto de visualizações. Isto é, por exemplo, a tarefa T09 que tem como questão "Existem *hotspots* no ano de 2012 e 2017?" implica o utilizador visualizar duas visualizações, um mapa de *hexabins* referente ao ano de 2012 e outro mapa de *hexabins* referente ao ano de 2017. As perguntas estão categorizadas em três categorias: Temática, Temporal e Espacial. Como os dados são espaço-temporais podemos fazer questões sobre dimensão temática dos dados, sobre a dimensão temporal ou sobre a dimensão espacial. Com esta categorização é possível agrupar questões que partilhem a mesma dimensão. Perguntas sobre características das ocorrências são categorizadas como "Temática", perguntas com componente temporal dos dados são categorizadas como "Temporal" e perguntas com componente espacial dos dados são categorizadas como "Espacial".

Tabela 4.1: Organização das perguntas de acordo com as ferramentas de deteção de padrões a serem avaliadas, as visualizações alvo e as questões aos utilizadores.

Foto Dashboard	Ferramenta	VIS	ID pergunta	Categoria	Questão
	Comparação das ocorrências nas áreas de influência dos quartéis	V2	P01	Temática	Qual a tipologia que apresenta maior número de ocorrências entre as áreas de influência?
		V4	P02	Temática	Qual das três áreas de influência em análise mobiliza, em média, mais elementos por ocorrência?
		V6	P03	Temporal	Qual das três áreas de influência em análise que em 2015 apresenta maior média de ocorrências diárias?
	Comparação das ocorrências nas freguesias de Lisboa	V2	P04	Temática	Qual a tipologia que apresenta maior número de ocorrências?
		V5	P05	Temática	Qual das três freguesias em análise mobiliza, em média, mais viaturas por ocorrência?
		V6	P06	Temporal	Qual das três freguesias em análise que em 2016 apresenta maior média de ocorrências diárias?
	Evolução espacial das tipologias das ocorrências ao longo dos anos	V1 e V6	P07	Espacial	Existem <i>hotspots</i> de ocorrências no ano de 2012 e 2017?
		V1, V6 e V8	P08	Espacial	Os <i>hotspots</i> entre os anos de 2012 e 2017 moveram-se?
	Análise do comportamento temporal das tipologias ao longo das estações do ano	V1 e V2	P09	Temporal	Qual o grupo que apresenta um pico de ocorrências na época do Verão?
		V4	P10	Temática	Qual é a tipologia do grupo que apresenta um pico de ocorrências na época de Verão?
		V3	P11	Espacial	As ocorrências do grupo que apresenta um pico no Verão estão concentradas numa zona de Lisboa ou estão dispersas?
	Análise do comportamento espacial das ocorrências	V1 e V2	P12	Espacial	Quais são os três grupos espaciais com mais ocorrências resultantes da aplicação de um algoritmo de agrupamento espacial às ocorrências de emergência?
		V2 e V3	P13	Temática	Qual é a tipologia mais comum para os grupos espaciais 0 e 11, respetivamente?
		V3	P14	Temática	Os grupos espaciais 0 e 11 partilham do mesmo comportamento em termos de tipos de tipologias?
		V6	P15	Temporal	Admitindo que o comportamento não difere de ano para ano, o grupo espacial 11 apresenta sazonalidade em termos mensais?

A avaliação utilizando esta metodologia secundária e devido a toda a situação que se vive neste tempo, a pandemia COVID-19, foi realizada em moldes remotos, não sendo necessário a presença física dos utilizadores para testes de avaliação experimental em laboratório. Esta abordagem trouxe alguns desafios, nomeadamente, o empenho do utilizador na experiência e a escolha de uma plataforma que possa ser acessível pelos utilizadores

Tabela 4.2: Organização das perguntas de acordo com as ferramentas de deteção de variáveis explicativas a serem avaliadas, as visualizações alvo e as questões aos utilizadores.

Foto Dashboard	Ferramenta	VIS	ID pergunta	Categoria	Questão
	Análise das ocorrências em redor de pontos de interesse em Lisboa	V2	P16	Temática	Qual é o nome do ponto de interesse que contém mais ocorrências em seu redor?
		V3	P17	Temporal	O número de ocorrências da subcategoria "Estações Ferroviárias", em termos gerais, aumentou, diminuiu ou manteve-se constante ao longo dos anos?
		V4	P18	Temática	Qual a tipologia mais comum em redor do ponto de interesse selecionado (Rossio)?
	Procura de variáveis explicativas para as ocorrências – Meteorologia	V1	P19	Temática	Em que condições de precipitação acontecem mais ocorrências?
		V2	P20	Temática	Quais os três identificadores das tipologias que contêm maior número de ocorrências sob chuva forte por ordem decrescente?
		V3	P21	Espacial	As ocorrências que acontecem sob chuva forte estão concentradas em alguma zona de Lisboa ou estão dispersas?
		V4	P22	Temática	Qual o maior número de elementos mobilizados para ocorrências que aconteceram sob chuva forte? E viaturas?
		V5	P23	Temporal	Considerando apenas as ocorrências que acontecem sob chuva forte, qual é o mês com maior número de ocorrências?

para a concretização do processo experimental.

Visto a realização desta avaliação experimental ser remota, houve a necessidade de agendar com os participantes reuniões remotas onde estive presente e que foram gravadas. A plataforma utilizada para as sessões virtuais e para a sua gravação foi o *Zoom*<sup>1</sup>. A minha presença nestas sessões serviu para dar instruções, acompanhar o processo de avaliação dos utilizadores, esclarecimento de dúvidas e recolha de algum *feedback*. Isto permite, para além de receber os dados referentes ao tempo de resposta e taxa de sucesso, receber outros comentários adicionais, como a existência de dúvidas ou dificuldades por parte dos utilizadores.

Relativamente às plataformas utilizadas, para a realização do questionário e recolha da métrica da taxa de sucesso foi utilizada a plataforma *Moodle* e para a recolha de informação sobre os tempos de resposta dos utilizadores foi utilizada a extensão *TimeYourWeb Time Tracker*<sup>2</sup> para o *browser Chrome*. A escolha da plataforma *Moodle* para a realização do questionário recaiu no facto de esta estar acessível a todos os alunos e pessoal da faculdade e permitir a criação de testes *online*. Para tal, foi criada uma disciplina onde os utilizadores foram inscritos e foi criado um teste que continha um questionário com todas as perguntas apresentadas nas tabelas 4.1. A plataforma *Moodle* permite a criação de diversos tipos de testes. Para este processo de avaliação experimental foi decidido que o questionário seria apresentado na forma de um teste de escolha múltipla. A decisão de apresentar um conjunto possível de resposta deve-se ao facto de que se disponibilizar um campo de texto livre os utilizadores irão responder de diversas formas às diferentes questões e utilizando um conjunto de respostas pré-definidas este problema irá ser combatido. Visto ser necessário a recolha da taxa de sucesso, o teste foi desenhado para que cada uma

<sup>1</sup><https://zoom.us/pt-pt/meetings.html>

<sup>2</sup><https://chrome.google.com/webstore/detail/timeyourweb-time-tracker/kfmlkgchpffnaphmlmjnimonlldbcphn>

das perguntas tenha uma e uma só resposta correta. O *Moodle* permite a possibilidade de baralhar as perguntas, isto é, baralhar a ordem pela qual as perguntas são apresentadas a diferentes utilizadores. Esta funcionalidade é interessante e foi utilizada em perguntas de alguns *dashboards*, nomeadamente perguntas dos *dashboards* de comparação das ocorrências nas áreas de influência dos quartéis, comparação das ocorrências nas freguesias de Lisboa, análise das ocorrências em redor de pontos de interesse em Lisboa e procura de variáveis explicativas para as ocorrências - Meteorologia. As perguntas dos restantes *dashboards* não foram baralhadas, pois as imagens de onde se obtém a resposta apresenta a resposta ou pistas para outras perguntas desse *dashboard*, sendo assim necessário haver ordem pela qual aparecem aos utilizadores. O teste também foi desenhado para apresentar cada pergunta numa página diferente. Deste modo, a utilização da extensão *TimeYourWeb Time Tracker* irá contar quanto tempo o utilizador demorou a responder à questão, visto que cada questão tem um URL diferente e a extensão contabiliza quanto tempo se passa em todos os URL que o utilizador visita.

Quanto à estrutura da avaliação experimental a figura 4.1 apresenta a sequência de passos da mesma. Como a figura 4.1 demonstra, inicialmente será apresentada uma introdução. Esta introdução não é extensa e serve para apresentar alguma contextualização da temática, isto é, apresenta alguns conceitos sobre as ocorrências do RSB e também apresenta a estrutura do questionário e instruções. De seguida, serão apresentadas duas perguntas de treino. Esta etapa apenas serve para o utilizador se familiarizar com o ambiente do questionário. Depois, serão apresentadas as várias perguntas presentes nas tabelas 4.1 e 4.2. Para a apresentação das perguntas será demonstrada a questão, uma figura do *dashboard* que está a ser avaliado e um conjunto possível de respostas à questão. Após as perguntas estarem todas respondidas, o utilizador será direccionado para um questionário demográfico, onde terá de preencher o campo de género, idade e se tem experiência em visualização de dados. Por fim, o utilizador é direccionado para a parte final da avaliação experimental onde irá encontrar um agradecimento.

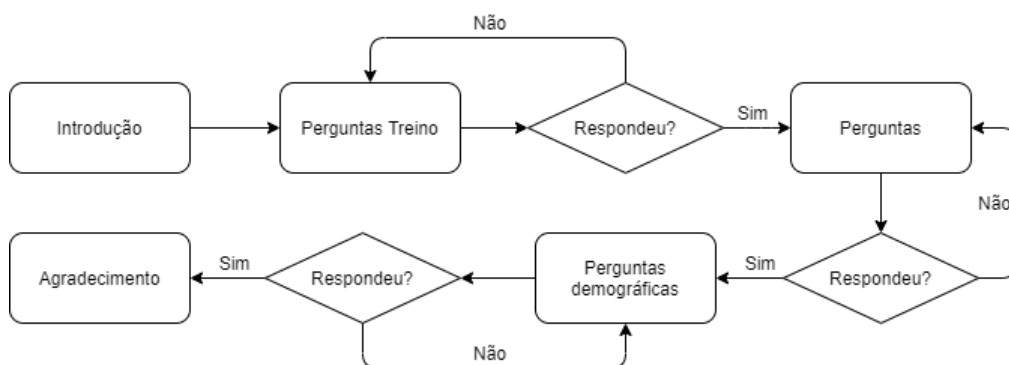


Figura 4.1: Fluxo da avaliação experimental.

Por fim, é importante salientar alguns aspetos sobre esta avaliação experimental. Devido à utilização de imagens a componente interativa das ferramentas desenvolvidas

deixa de existir. Outra nota a salientar sobre este processo de avaliação é que a leitura e interpretação da pergunta está contabilizada no tempo de resposta do utilizador.

## 4.2 Participantes

Todos os participantes deste processo de avaliação experimental são alunos e professores ligados à área de informática. No total participaram 24 voluntários, sendo que a maioria, 22 participantes, são colegas e alunos da faculdade (92%) e 2 são professores (8%). Todos os participantes, excepto os dois professores e dois colegas de curso, não tinham conhecimento dos resultados desta dissertação antes de participarem no processo de avaliação. Nenhum dos participantes teve explicação das visualizações antes da execução da avaliação experimental.

A figura 4.2 apresenta a distribuição da idade e género dos participantes. Como é mostrado na figura 4.2, a maioria dos participantes, cerca de 62%, tem idade entre 20 a 25 anos, de seguida, cerca de 26% dos participantes têm idade entre os 25 e 30 anos e as idades compreendidas entre 40-45, 45-50 e 50-55 apenas têm um participante cada. Como é visível, também pela figura 4.2, 58% dos participantes são do sexo masculino e 42% são do sexo feminino. Não existiram participantes femininas com idade acima de 30 anos.

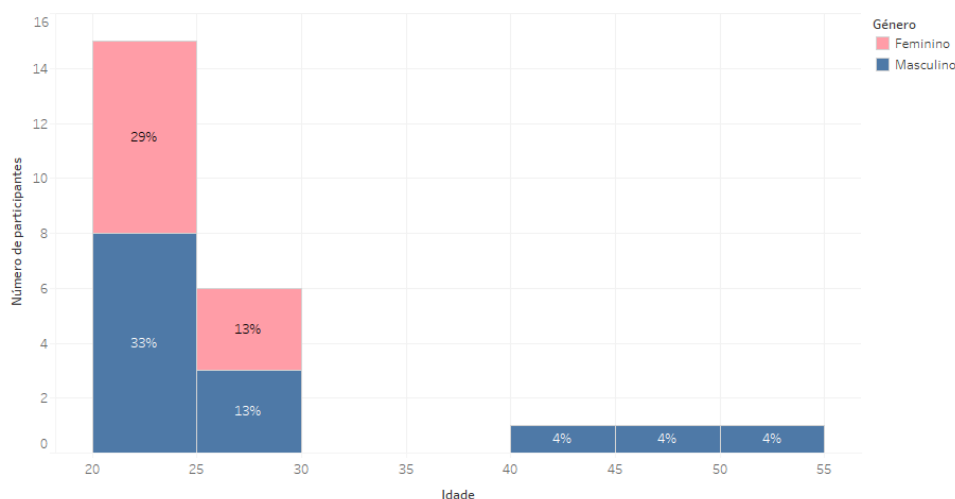


Figura 4.2: Dados demográficos dos participantes.

A figura 4.3 apresenta a experiência em visualização de dados dos participantes do processo de avaliação experimental. Como é mostrado na figura 4.3, a maioria dos participantes tem experiência em visualização de dados, 75%. Visto que a avaliação experimental foi feita maioritariamente por alunos da faculdade, quando os participantes afirmam que têm experiência em visualização de dados é porque estiveram inscritos na cadeira de Visualização Interactiva de Dados.

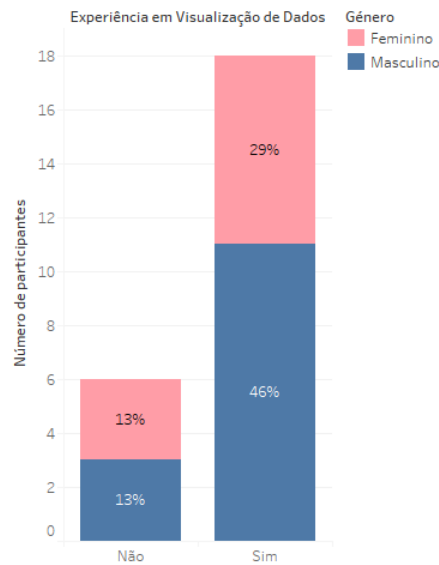


Figura 4.3: Experiência em visualização de dados dos participantes.

### 4.3 Resultados

Um dos principais objetivos da avaliação experimental foi a extração de dados sobre a taxa de sucesso das respostas às perguntas e o tempo de resposta dos participantes. A análise dos resultados foi feita recorrendo à taxa de sucesso e tempo de resposta dos utilizadores às perguntas. Para cada utilizador foram retiradas estas duas informações. Disto resultam 46 dados obtidos por cada participante, sendo que no total foram gerados 1104 dados. Foram também utilizadas as variáveis de género e experiência em visualização de dados dos utilizadores nesta análise de resultados.

De modo a apresentar os resultados obtidos, em termos da taxa de sucesso das respostas dadas pelos utilizadores, a figura 4.4 demonstra a percentagem de respostas corretas e erradas de cada pergunta presente no questionário. Como é visível pela figura 4.4 existem algumas perguntas cuja resposta foi sempre correta (P01, P03, P04, P07, P08, P09, P10, P12, P13, P14, P16, P17, P18 e P20). No entanto, existem questões em que isso não aconteceu (P02, P05, P06, P11, P15, P19, P21, P22 e P23). No caso das questões que têm respostas erradas por parte dos utilizadores há a salientar a pergunta P11, cuja categoria é espacial, onde houve mais participantes a errar (58%) do que a acertar (42%). As questões P6 (pergunta temporal) e P21 (pergunta espacial) têm, respectivamente, 29% e 25% de respostas erradas. Estas três perguntas, que contêm mais respostas erradas, serão analisadas mais à frente nesta seção. Também é visível na figura 4.4 que alguns *dashboards* tiveram uma taxa de sucesso de 100% nas respostas às perguntas que lhe estão associadas, o *dashboard* da evolução espacial das tipologias das ocorrências ao longo dos anos (*Hotspots*) e o da análise das ocorrências em redor de pontos de interesse em Lisboa (Pontos de Interesse). Os *dashboards* da comparação das ocorrências nas áreas de influência dos quartéis (Áreas de Influência), análise do comportamento temporal das tipologias ao longo das estações

do ano (Agrupamento Temporal) e análise do comportamento espacial das ocorrências (Agrupamento Espacial) apenas têm respostas erradas em uma das suas perguntas. Por fim, os *dashboards* da comparação das ocorrências nas freguesias de Lisboa (Freguesias) e da procura de variáveis explicativas para as ocorrências - Meteorologia (Meteorologia) têm apenas uma pergunta no seu conjunto de questões com uma taxa de sucesso de 100%, sendo que as restantes perguntas apresentam algumas respostas erradas.

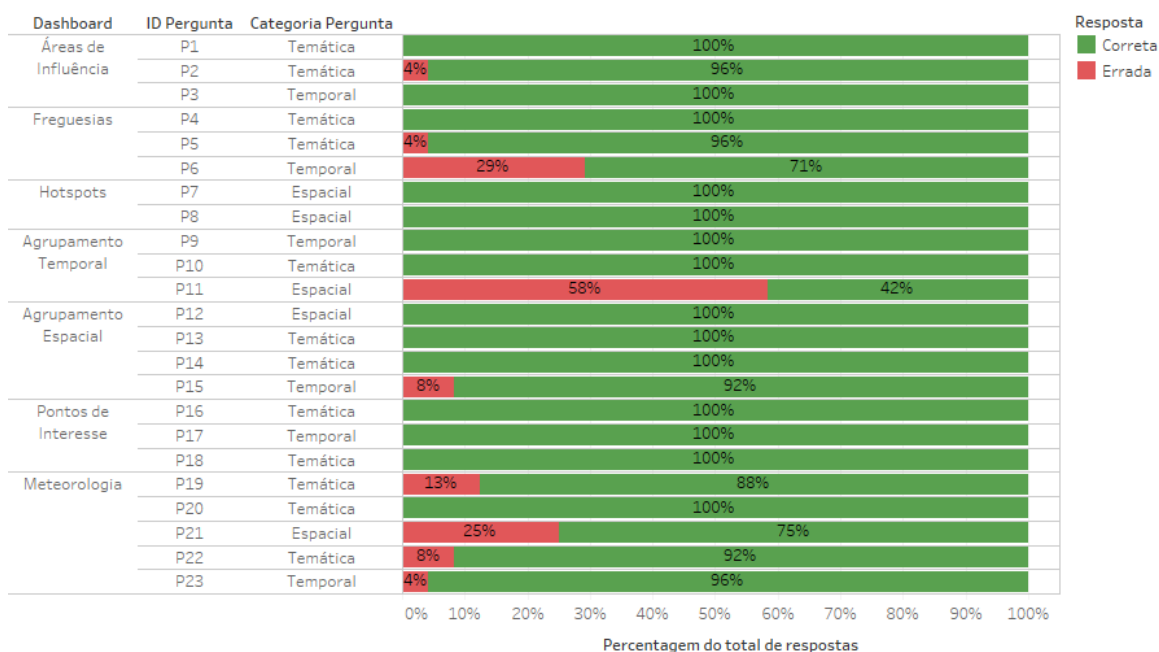


Figura 4.4: Quadro geral de acertos das perguntas.

Sendo que duas das perguntas com mais respostas erradas são da categoria espacial e uma da categoria temporal houve a necessidade de analisar a taxa de sucesso das respostas dos utilizadores pelas três categorias das perguntas, temática, temporal e espacial. A figura 4.5 apresenta a taxa de sucesso das três categorias das perguntas do questionário. Como a figura 4.5 demonstra a categoria que apresenta maior percentagem de respostas erradas por parte dos utilizadores é a categoria espacial seguida da temporal e, por fim, da temática. O resultado obtido na categoria espacial, em grande parte, deve-se aos resultados obtidos na pergunta P11, cuja percentagem de respostas erradas é superior às respostas corretas, e na pergunta P21.

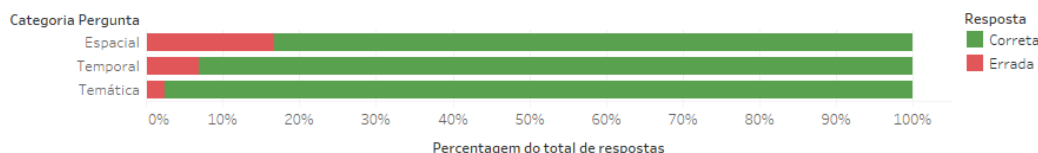


Figura 4.5: Taxa de sucesso das categorias das perguntas.

A variável de experiência do utilizador em visualização de dados é importante ser analisada em conjunto com a taxa de sucesso. A figura 4.6 apresenta a relação da taxa



de sucesso com a ausência, ou presença, de experiência em visualização de dados por parte dos participantes. Como é visível pela figura 4.6 não existe uma diferença que seja relevante entre a taxa de sucesso nas respostas às perguntas por parte de utilizadores com experiência e sem experiência em visualização de dados. Os participantes sem experiência apresentam 7% de respostas erradas e os participantes com experiências 8%. Isto é um dado positivo, pois demonstra que as visualizações são acessíveis a pessoas sem experiência em visualização de dados. Tanto os utilizadores com e sem experiência conseguem compreender as visualizações das ferramentas desenvolvidas. Portanto, com esta análise pode-se concluir que o factor de experiência em visualização de dados não é decisivo para a taxa de sucesso das respostas dos utilizadores.

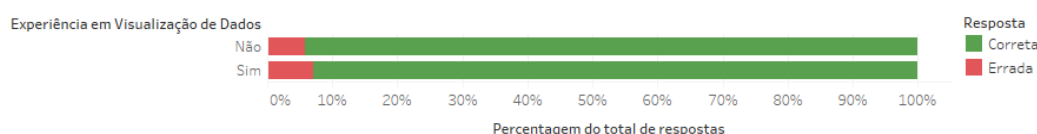


Figura 4.6: Taxa de sucesso das respostas dos utilizadores com e sem experiência em visualização de dados.

É necessário analisar os tempos de resposta para respostas corretas e incorretas. A figura 4.7 compara os resultados obtidos para a média do tempo de resposta para respostas corretas e erradas dos *dashboards* e tempo médio de resposta às perguntas dos *dashboards*. Analisando a figura 4.7 pode-se dizer que, na maioria dos *dashboards* (Áreas de Influência, Agrupamento Temporal, Agrupamento Espacial e Meteorologia), o tempo médio de resposta das respostas erradas é maior que o tempo médio de resposta em respostas corretas. Existe apenas um *dashboard* onde isto não se verifica, o *dashboard* das freguesias, onde, em média, os utilizadores demoram menos a responder erradamente do que corretamente. Existe um dado na figura 4.7 que salta à vista, o tempo médio de resposta para respostas erradas no *dashboard* do agrupamento espacial é muito superior ao tempo médio de resposta para respostas corretas. Este *dashboard* apenas apresenta respostas erradas na pergunta P15 (cerca de 8% de todas as respostas a essa pergunta), sendo este tempo médio relativo apenas à pergunta P15. A figura 4.7 apresenta também o tempo médio de resposta dos utilizadores às respostas de cada *dashboard*. A média do tempo de resposta dos *dashboards* estão mais próximos do tempo médio de resposta de respostas corretas do que erradas. De notar que, por exemplo, para o *dashboard* do agrupamento espacial o tempo médio de resposta do *dashboard* não é muito diferente do tempo médio de resposta das respostas corretas, porque houve poucos participantes a errar.

Após uma análise ao nível do *dashboard* é necessário fazer uma análise para a escala da imagem e da pergunta. Para tal, a visualização escolhida foi a caixa de bigodes para compreender melhor a distribuição do tempo de resposta obtido para cada participante neste processo de avaliação experimental. A figura 4.9 apresenta os tempos de resposta de cada participante para cada pergunta e identificador da imagem dos *dashboards*. Para responder às perguntas estas são acompanhadas por uma imagem do *dashboard* de onde

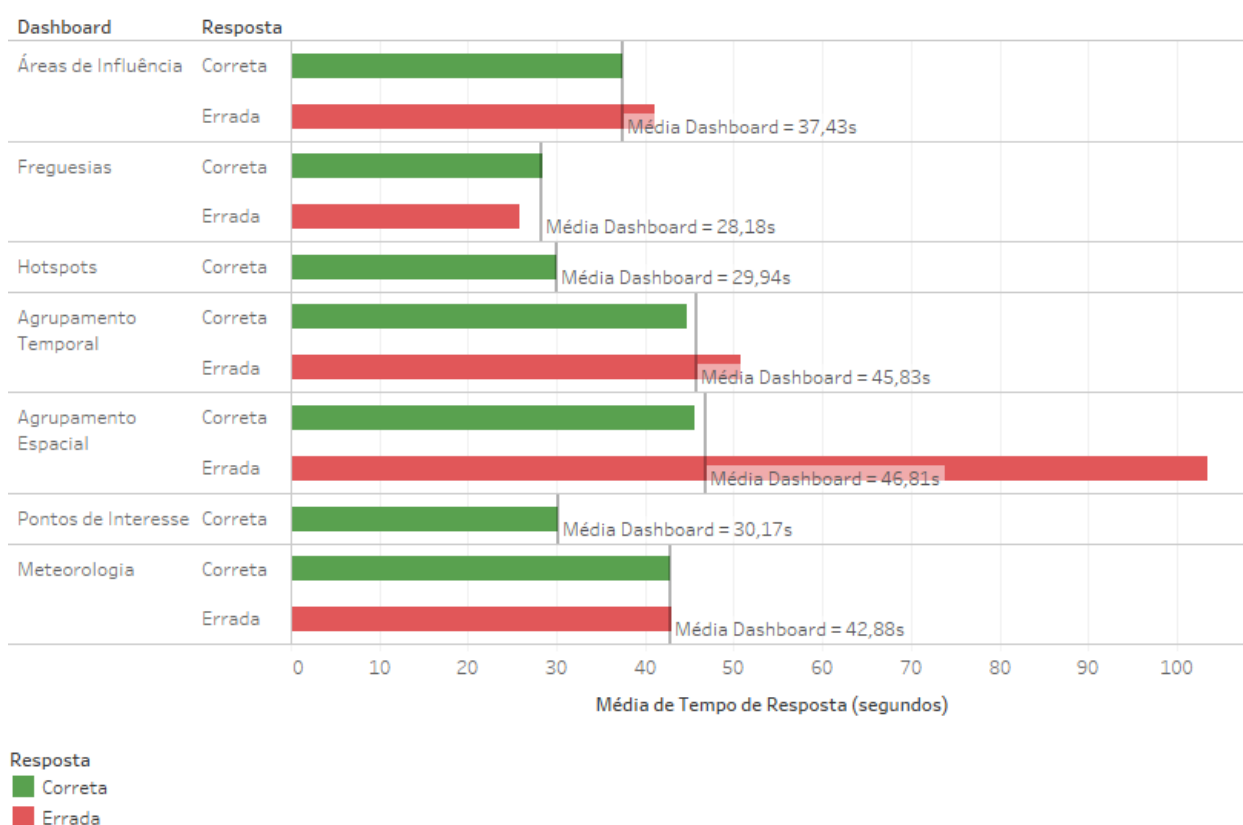


Figura 4.7: Média do tempo de resposta por *dashboard* para respostas corretas e erradas e tempo médio de resposta às perguntas dos *dashboards*.

se pode retirar a resposta, porém, dentro do mesmo *dashboard* as imagens podem ser diferentes. A figura 4.8 apresenta o exemplo do *dashboard* do agrupamento espacial que necessitou de duas imagens no questionário. A figura 4.8 apresenta o *dashboard* inicial e a figura 4.8a após a seleção de dois grupos espaciais, ou seja, de após uma interação com o mesmo. Uma questão pode ter resposta após uma interação com a ferramenta e como no processo de avaliação a componente de interação não existe é necessário apresentar ao participante uma imagem da ferramenta após a execução de uma ação. Devido a este facto existem *dashboards* que podem apenas apresentar a mesma imagem para obter a resposta a todas as perguntas correspondentes ou podem apresentar imagens do *dashboard* em diferentes estados para diferentes perguntas.

Como é possível observar pela figura 4.9 a distribuição dos tempos de resposta dos participantes varia de pergunta e *dashboard*. Os *dashboards* do agrupamento espacial e meteorologia são os que contém os valores do tempo de resposta mais elevados. Outro facto que se nota pela figura 4.9 é que o *dashboard* do agrupamento temporal apresenta alguma dispersão entre os valores máximos e mínimos nas perguntas associadas.

Analisar os tempos de resposta dos participantes por género é uma análise relevante. A figura 4.10 demonstra os tempos de resposta obtidos de cada participante na avaliação experimental por género. Como é visível pela figura 4.10, em média, as mulheres tendem



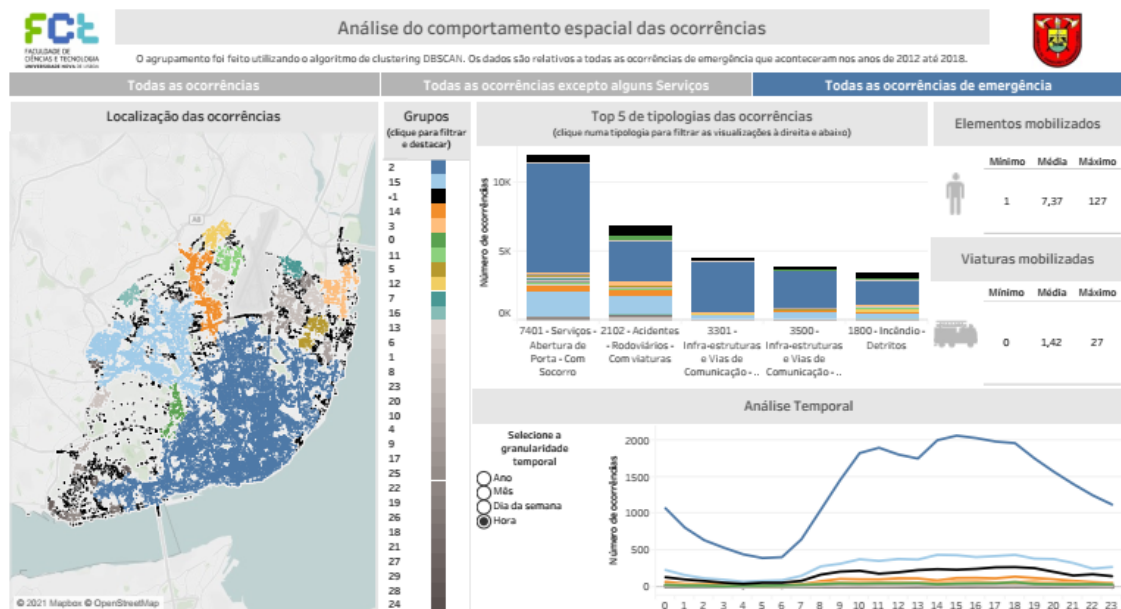
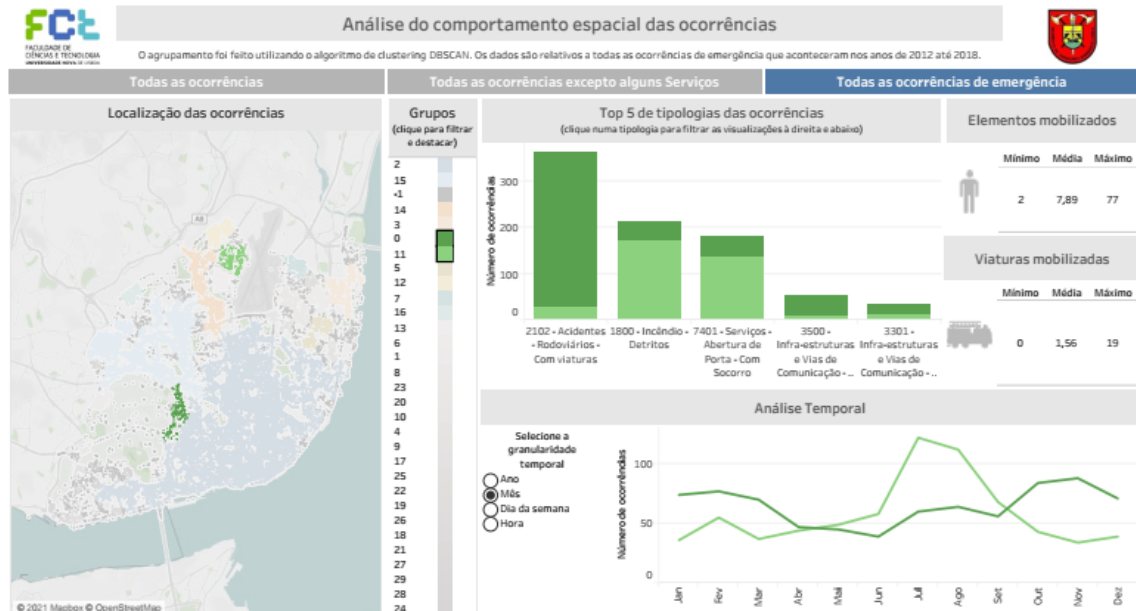
a Imagem inicial do *dashboard*.b Imagem do *dashboard* após uma ação.

Figura 4.8: Exemplo do *dashboard* de agrupamento espacial que necessitou de duas imagens para o questionário.

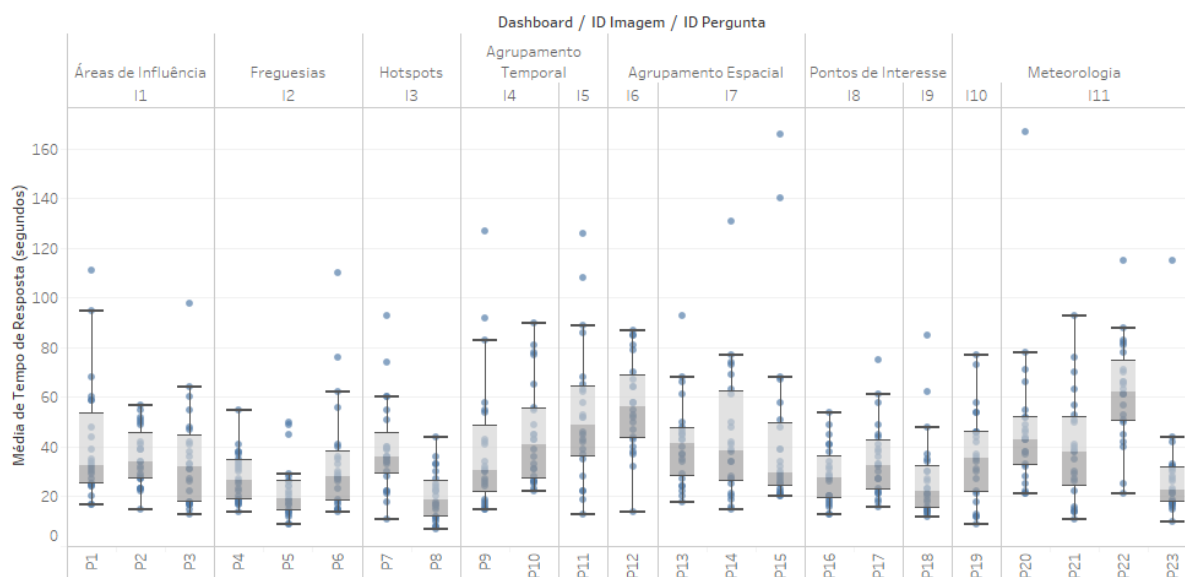


Figura 4.9: Tempo de resposta dos utilizadores por *dashboard*, identificador imagem e pergunta.

a ser mais rápidas a responder com uma média de 37,6 segundos e os homens um pouco mais lentos com uma média no tempo de respostas de 39,2 segundos. No entanto, a diferença entre as médias dos tempos de resposta das mulheres e dos homens não são muito diferentes, apenas diferem em 1,6 segundos. Outro aspecto a reter da figura 4.10 é que os tempos de resposta mais altos são relativos a homens, eles têm os cinco tempos de resposta mais elevados (167s, 166s, 140s, 131s e 127s). O tempo de resposta mais lento de uma mulher foi de 126 segundos.

Com os resultados anteriores, surge a curiosidade de fazer a mesma análise dos tempos de resposta dos participantes mas introduzindo a variável de experiência em visualização de dados em vez do género. A figura 4.11 apresenta os tempos de resposta obtidos de cada participante na avaliação experimental por experiência em visualização de dados. Observando a figura 4.11 é visível que a média do tempo de resposta para participantes sem e com experiência em visualização de dados é muito semelhante. Participantes sem experiência, em média, respondem em 38,8 segundos e participantes com experiência respondem, em média, 38,5 segundos. Portanto, olhando para valores médios, pessoas com experiência em visualização de dados tendem a ser mais rápidos a responder do que pessoas sem experiência. No entanto, a diferença entre as médias é de apenas 0,3 segundos, não sendo um valor relevante. Isto é, novamente, um indicativo positivo, visto que pessoas sem experiência conseguem responder mais ou menos de forma semelhante em termos de tempo de resposta dos que têm experiência, ou seja, este factor aparenta não ser decisivo. Outro aspeto que se pode observar na figura 4.11 é que os tempos de resposta mais elevados, pelo menos os dez mais altos, são relativos a participantes que têm experiência em visualização de dados. Como estive presente em todas as sessões da avaliação experimental posso dizer que os utilizadores com experiência tendem a demorar

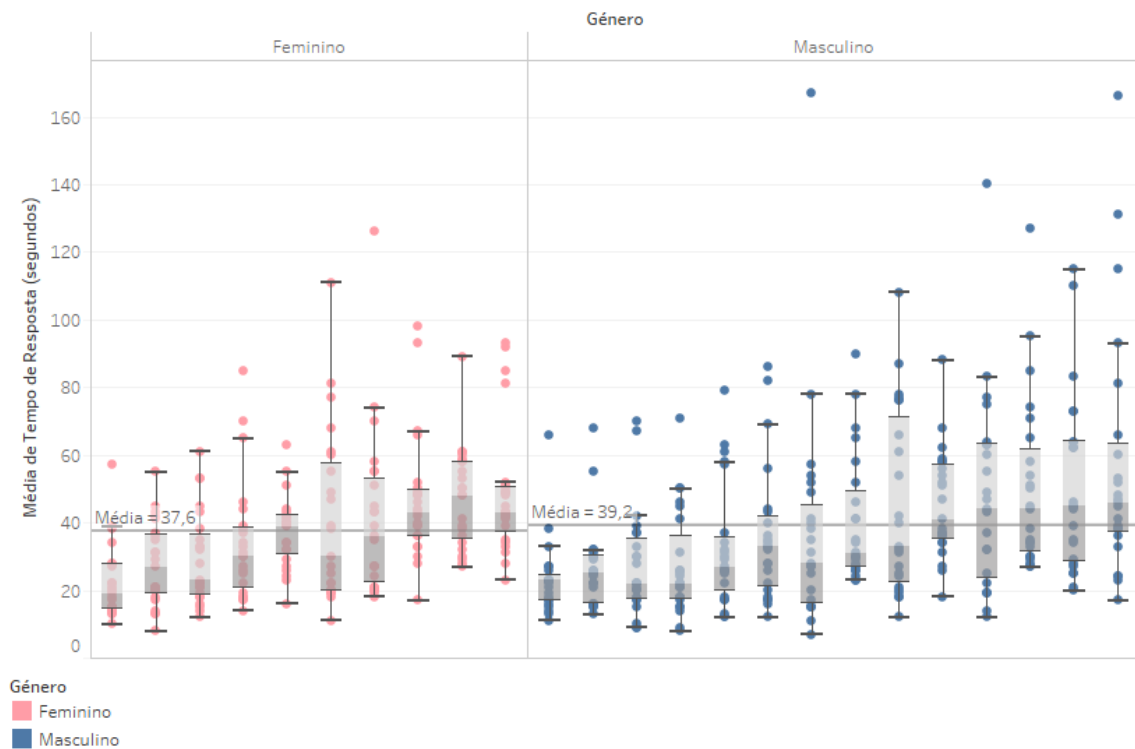


Figura 4.10: Tempo de resposta por utilizadores e por género.

mais tempo em certas perguntas, pois tendem a analisar mais as visualizações para obter respostas.

Cada *dashboard* apresenta um conjunto de perguntas. Em alguns dos *dashboards* foi utilizada a funcionalidade de baralhar as questões dessas ferramentas, ou seja, para diferentes participantes a ordem pela qual as perguntas apareciam no questionário era diferente. É importante analisar se a ordem pela qual as perguntas apareciam aos participantes é um factor decisivo relativamente ao tempo de resposta dos mesmos. A figura 4.12 apresenta a média dos tempos de resposta de cada pergunta dos diferentes *dashboards* pela ordem que aparecem no questionário dentro das perguntas do *dashboard*. Esta figura também apresenta o tempo médio de resposta das diversas perguntas. Analisando a figura 4.12 podemos observar que, à exceção de uma pergunta (a P21), em 13 perguntas a ordem pela qual aparecem no questionário parece influenciar o tempo de resposta. Isto quer dizer que as perguntas que aparecem em primeiro lugar no *dashboard* são as que demoram, em média, mais tempo a responder, isto porque sendo a primeira imagem do *dashboard* os participantes tendem a analisar quais são as visualizações da ferramenta e onde estão. A única pergunta onde isto não acontece é a questão P21 onde o tempo médio de resposta é mais elevado quando a pergunta aparece em terceiro, quarto e quinto lugar. Outra informação que podemos retirar da figura 4.12 é que existe uma aprendizagem no *dashboard* das freguesias. A primeira ferramenta apresentada no questionário é o *dashboard* das áreas de influência que é semelhante ao *dashboard* das freguesias, o *layout* e visualizações são iguais sendo que a única diferença é a organização administrativa

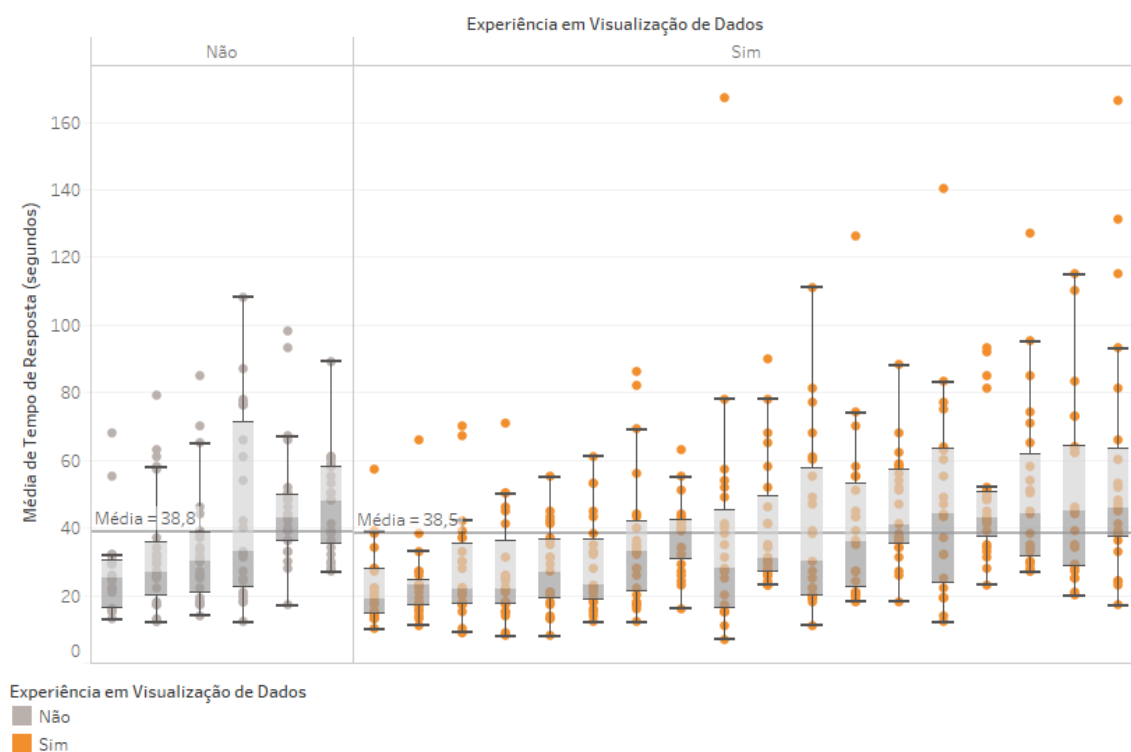


Figura 4.11: Tempo de resposta por utilizadores e por experiência em visualização de dados.

da área de Lisboa. Pode-se dizer que existe uma aprendizagem, pois o tempo médio de resposta das perguntas das freguesias diminuiu. A pergunta P1 corresponde à pergunta P04, a P02 à P05 e a P03 à P06. Assim, é visível que as perguntas P04 e P05 demoraram, em média, menos 10 segundos a ser respondidas que a P01 e P02. A pergunta P06, relativamente à P03, teve uma ligeira melhoria, cerca de 0,03 segundos.

Com a análise feita anteriormente é importante olhar não só para os tempos médios de resposta, mas para as distribuições do tempo de resposta dos participantes da avaliação experimental relativamente à ordem das perguntas. A figura 4.13 apresenta a distribuição do tempo de resposta dos diversos participantes relativamente às questões do *dashboard* das Áreas de Influência e da ordem pela qual aparecem aos participantes da avaliação experimental. Analisando a figura 4.13 pode-se observar que as distribuições do tempo de resposta dos participantes às diferentes questões nas diferentes ordens possíveis do questionário não são muito diferentes entre si, as suas diferenças não são muito relevantes, apesar de na média do tempo de resposta aparentar haver, maioritariamente, uma melhoria relevante da primeira para a segunda ordem no questionário. Portanto, a ordem pela qual as perguntas de um *dashboard* aparecem no questionário não aparenta ser um factor verdadeiramente influenciador do tempo de resposta.

Por fim, falta analisar as três questões que tiveram maior percentagem de respostas erradas, a P11, P06 e P21, respetivamente. A questão P11 é "As ocorrências do grupo que

### 4.3. RESULTADOS

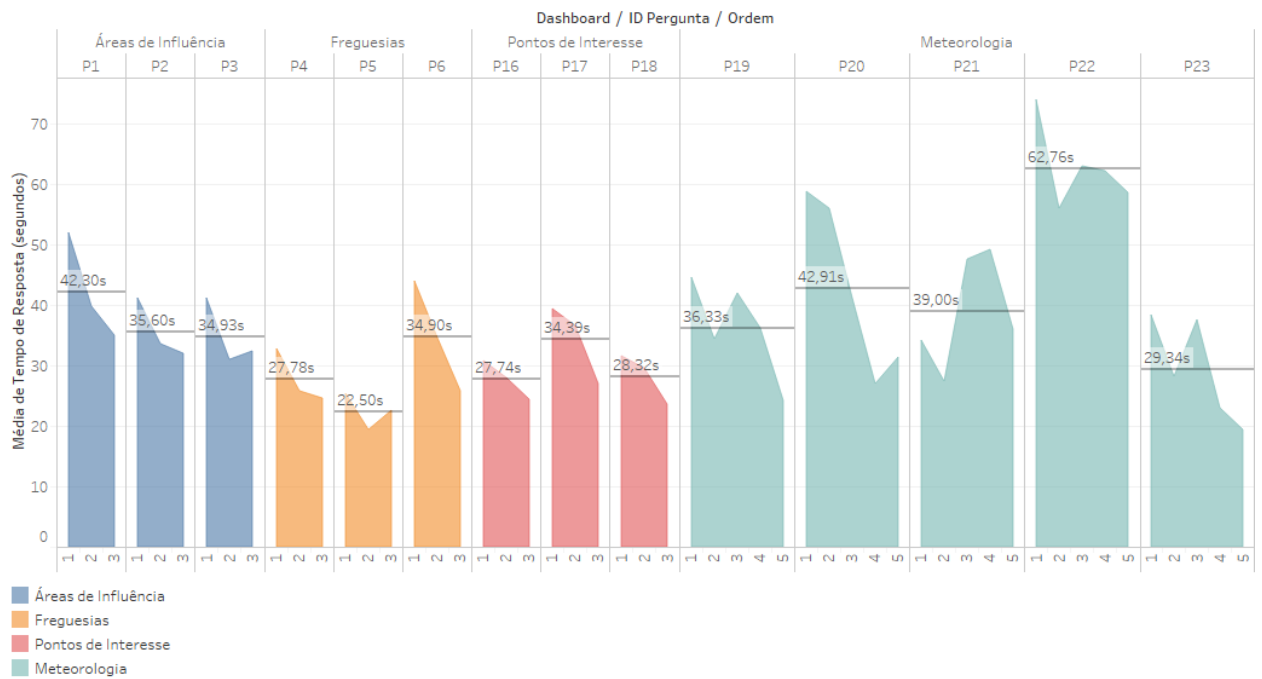


Figura 4.12: Média do tempo de resposta por *dashboard*, pergunta e ordem pela qual aparece no questionário.

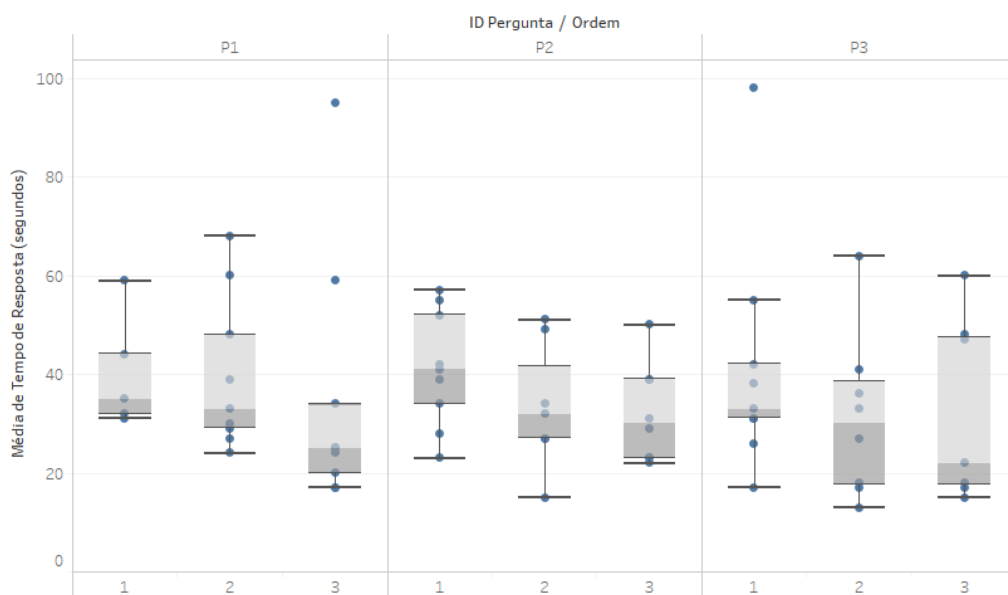


Figura 4.13: Tempo de resposta por utilizadores e pela ordem que aparecem no *dashboard* das Áreas de Influência.

apresenta um pico no Verão estão concentradas numa zona de Lisboa ou estão dispersas?" e pertence à categoria das perguntas espaciais. A figura 4.14 apresenta a visualização que continha a resposta a esta questão. Como a figura 4.14 demonstra a visualização alvo é um mapa de *hexabins*, talvez uma das visualizações criadas que os participantes menos têm conhecimento. Esta visualização não continha uma legenda de cores e tendo analisado as sessões dos participantes gravadas a principal dúvida nesta questão é de interpretação. Os participantes têm diferentes conceitos do que é concentrado numa zona ou disperso. A resposta considerada correta para esta questão era que as ocorrências estavam concentradas numa zona de Lisboa devido à existência de um hexágono muito vermelho na zona perto de Monsanto. No entanto, os participantes respondiam que as ocorrências estavam dispersas porque o concelho de Lisboa está quase todo preenchido por hexágonos claros ou porque existiam duas zonas mais a laranja que estão distantes. A questão P06 é "Qual das três freguesias em análise que em 2016 apresenta maior média de ocorrências diárias?" e pertence à categoria das perguntas temporais. A figura 4.15 apresenta a visualização que continha a resposta à questão P06. A resposta a esta questão encontra-se no gráfico de linhas e é a freguesia dos Olivais que apresenta maior média diária de ocorrências em 2016. Após análise das sessões com os participantes, estes erravam esta pergunta, pois liam na pergunta "média de ocorrências diárias" e olhavam para a visualização cujo título é "Média de ocorrências Diárias", presente no canto superior esquerdo da figura 4.15, que contém a média de todos os anos nos dados e não olhavam para o gráfico de linhas que contém a média por cada ano. Os participantes erraram esta questão porque respondiam Alvalade que é a freguesia que apresenta maior média de ocorrências diárias de todos os anos presentes nos dados. Portanto, o título da visualização influenciou negativamente a resposta a esta pergunta. A questão P21 é "As ocorrências que acontecem sob chuva forte estão concentradas em alguma zona de Lisboa ou estão dispersas?" e pertence à categoria espacial. Esta questão é semelhante à pergunta P11. A figura 4.16 apresenta a visualização que continha a resposta a esta questão e, novamente, esta visualização é um mapa de *hexabins*, tal como aconteceu com a questão P11. No entanto, esta visualização já contém uma legenda de cor. Novamente o conceito de concentrado e disperso de cada participante varia e daí os resultados serem menos positivos. A resposta considerada correta a esta questão é que as ocorrências estavam dispersas, pois existem vários hexágonos escuros dispersos pela região de Lisboa.

## 4.4 Conclusões

Com a execução do processo de avaliação experimental, apesar de não ter sido o inicialmente pensado, resultaram algumas conclusões.

Análises à variável género dos participantes demonstra que tanto os homens como as mulheres que participaram neste processo apresentam desempenhos semelhantes tanto no tempo que demoram a responder a uma pergunta como na taxa de sucesso da resposta.

A variável de experiência dos participantes em visualização de dados não é um factor



Figura 4.14: Visualização com resposta à pergunta P11.

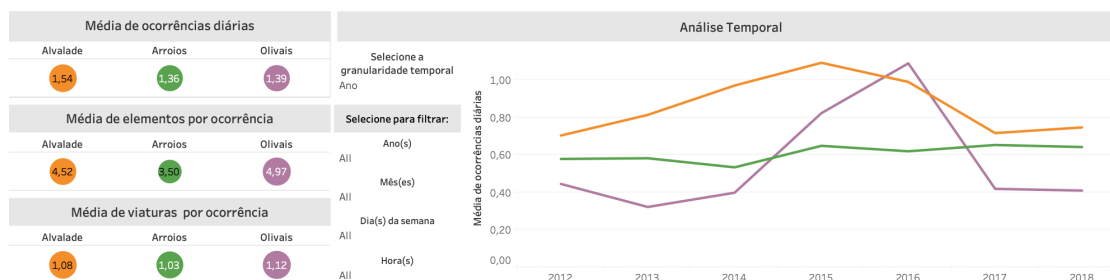


Figura 4.15: Visualização com resposta à pergunta P06.



Figura 4.16: Visualização com resposta à pergunta P21.

influenciador no tempo que os utilizadores demoram a responder às questões colocadas. Esta variável também não é um factor influenciador na taxa de sucesso de resposta às questões, ou seja, os participantes com experiência em visualização de dados não acertam mais perguntas que os participantes sem experiência. Ambos os tipos de participantes estão num nível de igualdade na taxa de sucesso. Este resultado é animador, pois é sinal que as visualizações são simples e fáceis de entender para todos os tipos de utilizadores, não requerendo conhecimento prévio.

Relativamente à ordem pela qual as perguntas dos *dashboards* aparecem no questionário, a ordem aparenta não ter relevância significativa no tempo que os utilizadores demoram a obter uma resposta às questões.

A taxa de sucesso foi menor em perguntas da categoria espacial, ou seja, questões com uma componente espacial foram as que obtiveram mais respostas erradas por parte dos participantes. As visualizações alvo das perguntas onde houveram alguns erros foram ambas em mapas de *hexabins*, uma visualização que, possivelmente, os participantes têm menos conhecimento prévio.

Para concluir, com a realização das análises feitas aos resultados obtidos no processo de avaliação experimental das ferramentas desenvolvidas surgiram algumas oportunidades de melhoramento nas mesmas, nomeadamente, mudanças de títulos de visualizações e colocação de legendas de cor.



## CONCLUSÕES E TRABALHO FUTURO

Com a realização desta dissertação foram cumpridos os objetivos principais do trabalho e ainda surgiram novos desenvolvimentos. Foi feito um levantamento das questões relevantes para esta dissertação em conjunto com a Câmara Municipal de Lisboa e o Regimento de Sapadores Bombeiros. Para cada uma das questões relevantes foi desenvolvido um *dashboard* que permite responder à questão em causa. Os *dashboards* desenvolvidos são ferramentas interativas de análise e exploração do conjunto de dados das ocorrências do RSB. Com as ferramentas resultantes desta dissertação espera-se que possam ajudar o comando operacional e estratégico do RSB a fazer a deteção de padrões temporais, espaciais e espaço-temporais nos seus dados das ocorrências e, ultimamente, o RSB possa decidir a alocação de meios de emergência baseada em evidências nos dados históricos das ocorrências. Relativamente a contributos secundários, desta dissertação resultou um relatório de qualidade dos dados das ocorrências do RSB, uma nova linguagem gráfica para a descrição de *dashboards* interativos e um novo conjunto de dados que reúne vários pontos de interesse, de diferentes categorias, do concelho de Lisboa.

Relativamente ao trabalho que poderá ser desenvolvido no futuro, visto não ter sido possível realizar a avaliação desejável das ferramentas desenvolvidas com utilizadores especializados na temática, considera-se que a realização desta avaliação com utilizadores experientes seja uma prioridade no futuro. É necessário validar as ferramentas criadas com pessoal do RSB apesar de os resultados obtidos na avaliação com utilizadores não experientes na temática terem sido positivos e daí terem surgido melhorias nas ferramentas desenvolvidas. Depois desta avaliação certamente surgirão possibilidades de novas melhorias e adição de funcionalidades aos *dashboards* interativos. Até poderão surgir novas ideias e um novo levantamento de questões relevantes que levam ao desenvolvimento de novas ferramentas visuais de análise e exploração de dados espaço-temporais. Outra questão que também poderá ser considerada futuramente, é a atualização do conjunto

de dados das ocorrências, isto é, este trabalho foi desenvolvido com dados do ano de 2012 até ao ano de 2018, sendo que, poderão adicionar, pelo menos, mais um ano de dados às ferramentas desenvolvidas. Para além da adição de novos anos nos dados, existe uma informação que é valiosa e fundamental para algumas análises, o tempo de resposta dos bombeiros a cada ocorrência. A adição desta informação possibilita análises sobre a distribuição e alocação dos vários elementos do [RSB](#).

## BIBLIOGRAFIA

- [1] N. Andrienko, G. Andrienko e P. Gatalsky. “Exploratory spatio-temporal visualization: an analytical review”. Em: *Journal of Visual Languages and Computing* 14.6 (2003). Visual Data Mining, pp. 503 –541. ISSN: 1045-926X. DOI: [https://doi.org/10.1016/S1045-926X\(03\)00046-6](https://doi.org/10.1016/S1045-926X(03)00046-6). URL: <http://www.sciencedirect.com/science/article/pii/S1045926X03000466>.
- [2] ArcGIS. *How Hot Spot Analysis (Getis-Ord Gi\*) works*. Acedido: 2020-01-31. URL: <https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm>.
- [3] ArcGIS. *What is a z-score? What is a p-value?* Acedido: 2020-01-31. URL: <https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/what-is-a-z-score-what-is-a-p-value.htm>.
- [4] A. Asgary, A. Ghaffari e J. Levy. “Spatial and temporal analyses of structural fire incidents and their causes: A case of Toronto, Canada”. Em: *Fire Safety Journal* 45.1 (2010), pp. 44 –57. ISSN: 0379-7112. DOI: <https://doi.org/10.1016/j.firesaf.2009.10.002>. URL: <http://www.sciencedirect.com/science/article/pii/S0379711209001313>.
- [5] D. Birant e A. Kut. “ST-DBSCAN: An algorithm for clustering spatial-temporal data”. Em: *Data & Knowledge Engineering* 60.1 (2007). Intelligent Data Mining, pp. 208 –221. ISSN: 0169-023X. DOI: <https://doi.org/10.1016/j.datak.2006.01.013>. URL: <http://www.sciencedirect.com/science/article/pii/S0169023X06000218>.
- [6] C. Brunsdon, J. Corcoran e G. Higgs. “Visualising space and time in crime patterns: A comparison of methods”. Em: *Computers, Environment and Urban Systems* 31.1 (2007). Extracting Information from Spatial Datasets, pp. 52 –75. ISSN: 0198-9715. DOI: <https://doi.org/10.1016/j.compenvurbsys.2005.07.009>. URL: <http://www.sciencedirect.com/science/article/pii/S0198971506000354>.
- [7] K. Cantwell, A. Morgans, K. Smith, M. Livingston, T. Spelman e P. Dietze. “Time of Day and Day of Week Trends in EMS Demand”. Em: *Prehospital Emergency Care* 19.3 (2015). PMID: 25664379, pp. 425–431. DOI: [10.3109/10903127.2014.995843](https://doi.org/10.3109/10903127.2014.995843). eprint: <https://doi.org/10.3109/10903127.2014.995843>. URL: <https://doi.org/10.3109/10903127.2014.995843>.

- [8] P. Chhetri, J. Corcoran e R. Stimson. “Exploring the spatio-temporal dynamics of fire incidence and the influence of socio-economic status: A case study from south east Queensland, Australia”. Em: *Journal of Spatial Science* 54.1 (2009), pp. 79–91. DOI: 10.1080/14498596.2009.9635168. eprint: <https://doi.org/10.1080/14498596.2009.9635168>. URL: <https://doi.org/10.1080/14498596.2009.9635168>.
- [9] A. Chohlas-Wood, A. Merali, W. Reed e T. Damoulas. “Mining 911 calls in New York City: temporal patterns, detection, and forecasting”. Em: *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence*. 2015.
- [10] S. W. Contributors. *Spyder*. Acedido: 2020-10-03. URL: <https://www.spyder-ide.org/>.
- [11] J. Corcoran, G. Higgs, C. Brunsdon e A. Ware. “The Use of Comaps to Explore the Spatial and Temporal Dynamics of Fire Incidents: A Case Study in South Wales, United Kingdom”. Em: *The Professional Geographer* 59.4 (2007), pp. 521–536. DOI: 10.1111/j.1467-9272.2007.00639.x. eprint: <https://doi.org/10.1111/j.1467-9272.2007.00639.x>. URL: <https://doi.org/10.1111/j.1467-9272.2007.00639.x>.
- [12] J. Corcoran, G. Higgs, C. Brunsdon, A. Ware e P. Norman. “The use of spatial analytical techniques to explore patterns of fire incidence: A South Wales case study”. Em: *Computers, Environment and Urban Systems* 31.6 (2007), pp. 623–647. ISSN: 0198-9715. DOI: <https://doi.org/10.1016/j.compenvurbsys.2007.01.002>. URL: <http://www.sciencedirect.com/science/article/pii/S0198971507000130>.
- [13] diagrams.net. *Diagram Software*. Acedido: 2020-10-03. URL: <https://www.diagrams.net/>.
- [14] L. Ertöz, M. Steinbach e V. Kumar. “Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data”. Em: *Proceedings of the 2003 SIAM international conference on data mining*. SIAM. 2003, pp. 47–58.
- [15] A. Espin. *Shared Nearest Neighbor Clustering Algorithm: Implementation and Evaluation*. Acedido: 2020-11-01. URL: <https://github.com/albert-espin/snn-clustering>.
- [16] M. Ester, H.-P. Kriegel, J. Sander, X. Xu et al. “A density-based algorithm for discovering clusters in large spatial databases with noise.” Em: *Kdd*. Vol. 96. 34. 1996, pp. 226–231.
- [17] P. Gatalaky, N. Andrienko e G. Andrienko. “Interactive analysis of event data using space-time cube”. Em: *Proceedings. Eighth International Conference on Information Visualisation, 2004. IV 2004*. (2004). DOI: 10.1109/iv.2004.1320137.

- 
- [18] IBM CRISP-DM. Acedido: 2020-01-15. URL: [https://www.ibm.com/support/knowledgecenter/SS3RA7\\_15.0.0/com.ibm.spss.crispdm.help/crisp\\_overview.htm](https://www.ibm.com/support/knowledgecenter/SS3RA7_15.0.0/com.ibm.spss.crispdm.help/crisp_overview.htm).
- [19] A. Inc. *Anaconda*. Acedido: 2020-10-03. URL: <https://www.anaconda.com/products/individual>.
- [20] A. K. Jain. "Data clustering: 50 years beyond K-means". Em: *Pattern Recognition Letters* 31.8 (2010). Award winning papers from the 19th International Conference on Pattern Recognition (ICPR), pp. 651–666. ISSN: 0167-8655. DOI: <https://doi.org/10.1016/j.patrec.2009.09.011>. URL: <http://www.sciencedirect.com/science/article/pii/S0167865509002323>.
- [21] H. L. Jasso, W. S. Hodgkiss, C. Baru, T. Fountain, D. Reich e K. Warner. "Spatiotemporal characteristics of 9-1-1 emergency call hotspots". Em: 2007.
- [22] C. R. Jennings. "Socioeconomic Characteristics and Their Relationship to Fire Incidence: A Review of the Literature". Em: *Fire Technology* 35.1 (1999), pp. 7–34. ISSN: 1572-8099. DOI: [10.1023/A:1015330931387](https://doi.org/10.1023/A:1015330931387). URL: <https://doi.org/10.1023/A:1015330931387>.
- [23] N. Lavrac, D. Jesenovec, N. Trdin e N. M. Kosta. "Mining spatio-temporal data of traffic accidents and spatial pattern visualization". Em: *Metodoloski zvezki* 5.1 (2008), p. 45.
- [24] S. L. Linton, J. M. Jennings, C. A. Latkin, M. B. Gomez e S. H. Mehta. "Application of Space-Time Scan Statistics to Describe Geographic and Temporal Clustering of Visible Drug Activity". Em: *Journal of Urban Health* 91.5 (2014), pp. 940–956. ISSN: 1468-2869. DOI: [10.1007/s11524-014-9890-7](https://doi.org/10.1007/s11524-014-9890-7). URL: <https://doi.org/10.1007/s11524-014-9890-7>.
- [25] C. M. de Lisboa. *Segurança e Prevenção*. Acedido: 2020-01-26. URL: <https://www.lisboa.pt/cidade/seguranca-e-prevencao/regimento-de-sapadores-bombeiros>.
- [26] Q. Liu, M. Deng, J. Bi e W. Yang. "A novel method for discovering spatio-temporal clusters of different sizes, shapes, and densities in the presence of noise". Em: *International Journal of Digital Earth* 7.2 (2014), pp. 138–157. DOI: [10.1080/17538947.2012.655256](https://doi.org/10.1080/17538947.2012.655256). eprint: <https://doi.org/10.1080/17538947.2012.655256>. URL: <https://doi.org/10.1080/17538947.2012.655256>.
- [27] *LxDataLab*. Acedido: 2020-01-15. URL: <https://lisboainteligente.cm-lisboa.pt/lxdatalab/>.
- [28] N. Malleson e M. A. Andresen. "Spatio-temporal crime hotspots and the ambient population". Em: *Crime Science* 4.1 (2015), p. 10. ISSN: 2193-7680. DOI: [10.1186/s40163-015-0023-8](https://doi.org/10.1186/s40163-015-0023-8). URL: <https://doi.org/10.1186/s40163-015-0023-8>.

- [29] T. Nakaya e K. Yano. “Visualising Crime Clusters in a Space-time Cube: An Exploratory Data-analysis Approach Using Space-time Kernel Density Estimation and Scan Statistics”. Em: *Transactions in GIS* 14.3 (2010), pp. 223–239. DOI: [10.1111/j.1467-9671.2010.01194.x](https://doi.org/10.1111/j.1467-9671.2010.01194.x). eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9671.2010.01194.x>. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9671.2010.01194.x>.
- [30] S. V. Nath. “Crime Pattern Detection Using Data Mining”. Em: *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. WI-IATW '06. USA: IEEE Computer Society, 2006, 41–44. ISBN: 0769527493. DOI: [10.1109/WI-IATW.2006.55](https://doi.org/10.1109/WI-IATW.2006.55). URL: <https://doi.org/10.1109/WI-IATW.2006.55>.
- [31] S. Notícias. *Sapadores Bombeiros de Lisboa contabilizam 126 ocorrências devido à tempestade de granizo*. Acedido: 2020-01-30. URL: <https://sicnoticias.pt/pais/2011-04-29-sapadores-bombeiros-de-lisboa-contabilizam-126-ocorrencias-devido-a-tempestade-de-granizo>.
- [32] QGIS. *Discover QGIS*. Acedido: 2020-10-03. URL: <https://qgis.org/en/site/about/index.html>.
- [33] S. Rodrigues e A. Figueiras. “There and then: interacting with spatio-temporal visualization”. Em: ().
- [34] M. Y. Santos, J. M. Pires, G. Moreira, R. Oliveira, F. Mendes e C. Costa. “Geo-spatial analytics using the dynamic ST-SNN Approach”. Em: *Lecture Notes in Engineering and Computer Science* 1 (2015), pp. 285–290.
- [35] A. Selvam, B. Thiagarajan e T. K. Thivakaran. “Mining Patterns from 9-1-1 Calls Dataset”. Em: 2016. DOI: [10.5120/ijais2016451632](https://doi.org/10.5120/ijais2016451632).
- [36] T. SOFTWARE. *Tableau Desktop*. Acedido: 2020-10-03. URL: <https://www.tableau.com/products/desktop>.
- [37] T. SOFTWARE. *Tableau Online*. Acedido: 2020-10-03. URL: <https://www.tableau.com/products/cloud-bi>.
- [38] T. SOFTWARE. *Tableau Prep*. Acedido: 2020-10-03. URL: <https://www.tableau.com/learn/get-started/prep>.
- [39] H. F. Tork. “Spatio-temporal clustering methods classification”. Em: *Doctoral Symposium on Informatics Engineering*. Vol. 1. 1. Faculdade de Engenharia da Universidade do Porto Porto, Portugal. 2012, pp. 199–209. DOI: [10.13140/RG.2.1.3812.7204](https://doi.org/10.13140/RG.2.1.3812.7204).

- [40] K. Wuschke, J. Clare e L. Garis. "Temporal and geographic clustering of residential structure fires: A theoretical platform for targeted fire prevention". Em: *Fire Safety Journal* 62 (2013). Special Issue on Spatial Analytical Approaches in Urban Fire Management, pp. 3 –12. ISSN: 0379-7112. DOI: <https://doi.org/10.1016/j.firesaf.2013.07.003>. URL: <http://www.sciencedirect.com/science/article/pii/S0379711213001148>.







## RELATÓRIO DE QUALIDADE DOS DADOS

A qualidade dos dados é importante para satisfazer os requisitos do uso pretendido. Ao trabalhar com dados de alta qualidade, o processamento e a análise dos mesmos torna-se mais fácil. Segundo a norma internacional ISO 9000:2015, algumas características da qualidade dos dados são a completude, validade, precisão, consistência, disponibilidade e actualidade.

Este relatório irá avaliar a qualidade dos dados das ocorrências do Regimento de Sapadores Bombeiros disponibilizados pela Câmara Municipal de Lisboa. Os principais objectivos da elaboração deste relatório são entender e explorar os dados e, consequentemente, analisar a sua qualidade.

A metodologia usada foi baseada numa fase do modelo do processo de prospecção de dados do *CRISP-DM* [18], a fase de compreender os dados. As etapas desta fase, e consequentemente, deste relatório são as seguintes:

- Reunir dados iniciais;
- Descrição dos dados;
- Exploração dos dados;
- Verificação da qualidade dos dados.

Este relatório poderá ser utilizado para a seguinte fase do modelo do processo de prospecção de dados, a fase da preparação dos dados. Esta fase é onde ocorre a tomada de decisões sobre limpeza e tratamento dos dados.

## A.1 Reunir dados iniciais

Os dados referentes às ocorrências do Regimento de Sapadores Bombeiros, que foram disponibilizados pela Câmara Municipal de Lisboa, estão presentes num ficheiro Excel. Este ficheiro está separado em três folhas. As folhas apresentam os seguintes números de linhas:

- 1ª folha contém 65 000 linhas;
- 2ª folha contém 65 000 linhas;
- 3ª folha contém 5 201 linhas.

De notar que os números de linhas acima indicados contam com o cabeçalho.

As diferentes folhas apresentam o mesmo número de colunas, 25, e as colunas são todas semelhantes entre as diferentes folhas. Portanto, não existirá problemas em fundir os dados das folhas.

Foram disponibilizados metadados para este conjunto de dados. A tabela A.1 sintetiza esses metadados.

Ao analisar a tabela A.1, conclui-se que estes dados apresentam componentes espaciais e temporais.

A coluna referente ao código identificador do sistema de ocorrências do Regimento de Sapadores Bombeiros não aparenta ser um atributo interessante para análise, pois é um identificador privado de um sistema à parte. Este atributo não será alvo de análise neste relatório.

## A.2 Descrição dos dados

Os dados sobre as ocorrências do Regimento de Sapadores Bombeiros contêm, no total, 135 198 registos de ocorrências e 24 colunas de interesse.

A tabela A.2 apresenta a descrição dos tipos de valores dos atributos presentes nos dados, de acordo com os metadados disponibilizados, classificando-os como numéricos ou simbólicos e dando alguns exemplos de possíveis valores que são encontrados nestes atributos.

Como se pode verificar pela tabela A.1, a maioria dos tipos de valor nestes dados são numéricos, devido aos identificadores e aos códigos SIG que fazem a identificação do local das ocorrências. Os atributos simbólicos, na sua maioria, dizem respeito a características das próprias ocorrências, como por exemplo, a sua tipologia. Também se encontram dois atributos temporais, a data da ocorrência e a data de alteração do estado da ocorrência. Existem apenas dois atributos que são booleanos, o identificador de suspensão antes da activação de meios e o identificador de falso alarme após a activação de meios.

Dois atributos apresentam esquemas de codificação. A descrição do estado de ocorrência apresenta três valores possíveis, "FA", "FO" e "A". O valor "FA" significa fecho

Tabela A.1: Significado das colunas dos dados

Atributo	Descrição
ID	Identificador único da ocorrência
Data e hora	Data e hora do registo da ocorrência
Longitude	Longitude da ocorrência
Latitude	Latitude da ocorrência
Código de via	Código de via do local da ocorrência
Código SIG	Código SIG do edifício ou código de via da ocorrência
Código SIG da fracção	Código SIG da fracção da ocorrência
Complemento do local	Detalhes sobre o local da ocorrência
Designação do local	Morada da ocorrência
Código SIG da freguesia	Código SIG da freguesia onde a ocorrência aconteceu
Descrição da freguesia	Nome da freguesia onde a ocorrência aconteceu
Data e hora alteração do estado	Data e hora da alteração do estado da ocorrência
ID do estado	Identificador do estado da ocorrência
Descrição do estado	Descrição do estado da ocorrência
ID da tipologia	Identificador da tipologia da ocorrência
Descrição da tipologia	Descrição da tipologia da ocorrência
Descrição da ocorrências	Detalhes sobre a ocorrência
ID da entidade	Identificador da entidade que iniciou a ocorrência
Descrição da entidade	Sigla da entidade que iniciou a ocorrência
Indicador do grau de risco	Grau de risco da ocorrência
Indicador de suspensão	Indica se a ocorrência foi suspensa ou não
Identificador de falso alarme	Indica se a ocorrência foi falso alarme depois da activação de meios ou não
ID sistema de ocorrências do RSB	Código identificador do sistema de ocorrências do RSB
Número de viaturas	Número de viaturas a enviar para o local da ocorrência
Estimativa número de elementos	Estimativa do número de elementos que serão enviados para o local da ocorrência

administrativo, "FO" significa fecho operacional e "A" significa accionada. O outro atributo que apresenta codificação é o indicador do risco. Este atributo tem dois valores numéricos possíveis, 1 e 2. O valor 1 codifica para um grau de risco elevado e o valor 2 para um grau de risco moderado.

### A.3 Exploração dos dados

Para conhecer os dados é necessário estes serem explorados. Esta fase é essencial para a familiarização com os dados e para a análise da qualidade dos mesmos. A exploração irá ser efectuada através de técnicas de visualização de dados, que irão permitir detectar problemas neste conjunto de dados, como por exemplo, problemas de consistência e de completude.

A exploração dos dados irá permitir que seja definido um conjunto de tarefas de limpeza e transformação de dados, que poderão ocorrer numa fase posterior, de modo a

Tabela A.2: Tipos de valor dos atributos

Atributo	Tipo de valores	Exemplo Valores possíveis
ID	Numérico	338 210 032 015
Data e hora	Simbólico	10/03/2015 15:45:00
Longitude	Numérico	-9,1447032488817
Latitude	Numérico	38,7591857636647
Código de via	Numérico	66 669
Código SIG	Numérico	4200102001001
Código SIG da fração	Numérico	0100306005001001
Complemento do local	Simbólico	Lote 49 2º eq
Designação do local	Simbólico	Avenida do Brasil - 53 (Pav 35)
Código SIG da freguesia	Numérico	103
Descrição da freguesia	Simbólico	Alvalade
ID do estado	Numérico	4, 3, 2
Data e hora alteração do estado	Simbólico	15/01/2018 14:53:52
Descrição do estado	Simbólico	FA, FO e A
ID da tipologia	Numérico	11
Descrição da tipologia	Simbólico	1300 - Incêndio - Inculto
Descrição da ocorrência	Simbólico	Masculino 85 anos CODU 3504
ID da entidade	Numérico	1
Descrição da entidade	Simbólico	PUB
Indicador do grau de risco	Numérico	1, 2
Indicador de suspensão	Boleano	N, S
Identificador de falso alarme	Boleano	0, 1
Número de viaturas	Numérico	1, 2
Estimativa número de elementos	Numérico	6

aumentar a qualidade dos dados.

### A.3.1 Identificador da ocorrência

Ao explorar os dados, relativamente aos identificadores que referem cada ocorrência, verificou-se que estes identificadores são únicos, isto é, não foram detectados duplicados. No total existem 135 198 identificadores, ou seja, existem exactamente 135 198 ocorrências registadas nos dados do Regimento de Sapadores Bombeiros.

### A.3.2 Data de ocorrência

Ao analisar o atributo referente à data das ocorrências, verificou-se que os dados dizem respeito aos anos de 2011 até 2018. O primeiro dia de registo de ocorrências é 17 de Janeiro de 2011 e o último dia de registo é 17 de Dezembro de 2018.

A figura A.1 apresenta a evolução do número de ocorrências atendidas pelo Regimento de Bombeiros Sapadores entre os anos de 2011 e 2018. A figura A.1 revela que o ano de 2011 apresenta poucas ocorrências relativamente aos outros anos. Apenas existe registo de 2 138 ocorrências em 2011. Analisando este ano, mas na escala de meses, verifica-se

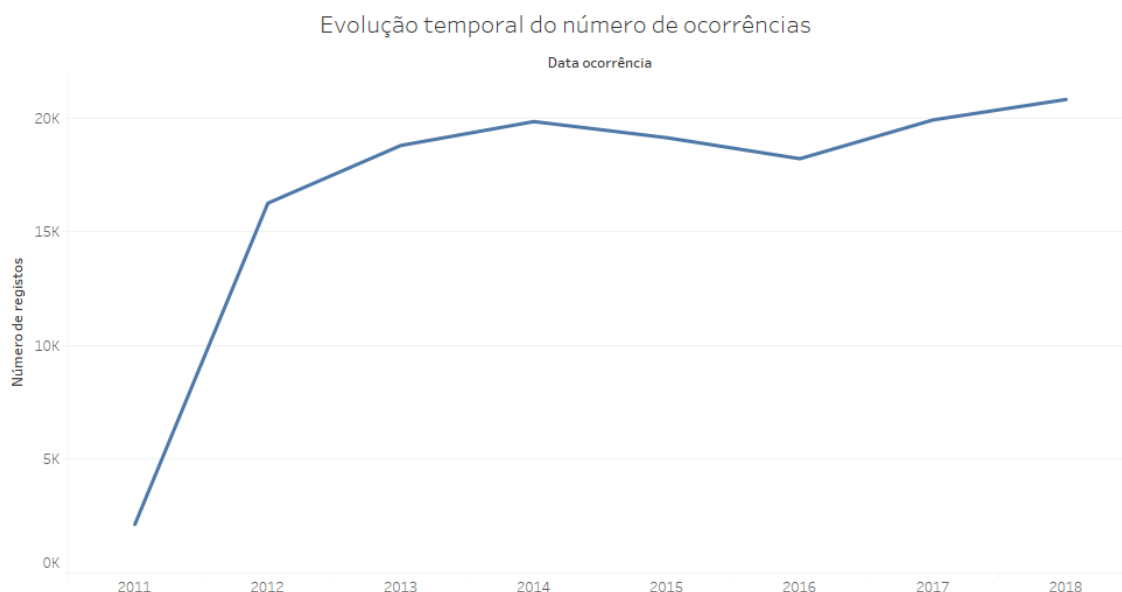


Figura A.1: Evolução temporal, ao longo dos anos, do número de ocorrências.

que existem meses com poucas ocorrências ou sem elas. O mês de Março, Junho e Julho não tem ocorrências registradas, existe apenas uma ocorrência registrada nos meses de Janeiro, Abril e Maio e três ocorrências no mês de Fevereiro. Este dados não espelham a realidade é muito pouco provável que não sejam registradas ocorrências num período de um mês.

Os anos de 2012 até 2018 não diferem muito significativamente no número total de ocorrências registradas. Analisando estes anos, mas na escala de meses, não são encontradas anomalias. Em todos os meses existem várias ocorrências registradas ao longo dos anos.

A exploração deste atributo revelou problemas no ano de 2011, existem meses que não registam ocorrências, não sendo um espelho do que acontece na realidade. A notícia [31] vem a confirmar que os dados das ocorrências de 2011 não espelham a realidade. Esta notícia faz referência à existência de 126 ocorrências no mês de Abril devido a uma tempestade de granizo. Nos dados, o mês de Abril apenas tem registo de uma ocorrência.

### A.3.3 Latitude e longitude

Ao explorar os dados relativamente à existência de coordenadas geográficas, latitude e longitude, verificou-se que existem ocorrências que não contêm estas informações.

A figura A.2 apresenta o número de ocorrências, do Regimento de Sapadores Bombeiros, que contêm coordenadas geográficas e o número de ocorrências que não as contêm. Como demonstra a figura A.2, 1.22% das ocorrências não apresentam coordenadas geográficas. No entanto, algumas dessas ocorrências, apresentam códigos de via ou códigos SIG que identificam o local da ocorrência.

A figura A.3 apresenta a distribuição das ocorrências, do Regimento de Sapadores



Figura A.2: Número de ocorrências com e sem dados de latitude e longitude.

### Distribuição das ocorrências sem coordenadas geográficas ao longo dos anos

Coordenadas	Data ocorrência					Total geral	% total de nº registos	
	2011	2012	2015	2016	2017			
Nulo	46,26%	51,13%	0,06%	0,06%	2,50%	100,00%	0,06%	51,13%

Figura A.3: Distribuição das ocorrências sem latitude e longitude ao longo dos anos.

Bombeiros, que não apresentam informações sobre a latitude e longitude do local da ocorrência, ao longo dos anos. Como a figura A.3 demonstra, os anos que contêm mais nulos nas coordenadas são os anos de 2011 e 2012 com 46,26% e 51,13% do total de ocorrências com as coordenadas a nulo, respectivamente. Os anos de 2015, 2016 e 2017 também apresentam algumas falhas neste atributo mas não são muito significativas.

Foi feita uma análise mais profunda aos dados que contêm coordenadas geográficas para validar os valores de latitude e longitude. Esta análise permitiu detectar a presença de duas ocorrências com erros nos valores da latitude e/ou longitude. Uma destas ocorrências apresenta coordenadas que a colocam no meio do Mar Mediterrâneo e a outra na Holanda.

Com a exploração destes atributos foi possível detectar, novamente, problemas nos dados de 2011. Para além do ano de 2011 não ser um ano representativo da realidade apresenta, também, problemas com a ausência de coordenadas geográficas. O ano de 2012 revelou ser o ano com maior número de ocorrências sem informações sobre a latitude e longitude. De notar que algumas ocorrências com este problema apresentam outros atributos que são capazes de derivar as coordenadas. Apenas foram detectadas duas ocorrências que apresentam a latitude e/ou longitude erradas.

#### A.3.4 Morada

Existem vários atributos que são relativos à morada das ocorrências, o código da via, descrição do local, código SIG, código SIG da fracção e complemento do local.

O código da via é um código que representa uma rua, neste caso a rua do local da ocorrência. Este atributo está correlacionado com o atributo da descrição do local, isto é, a morada da ocorrência. No entanto, o atributo de descrição da morada do local, por vezes, apresenta coordenadas geográficas em vez da morada.

Ao analisar melhor os valores do atributo código da via, foram detectados códigos que aparentam ser de locais desconhecidos ou sem morada, por exemplo, códigos começados por 9999999 e o código 0.



Figura A.4: Número de ocorrências com e sem morada.

A figura A.4 mostra o número de ocorrências, do Regimento de Sapadores Bombeiros, que têm o atributo da descrição do local preenchido com a morada e as que têm preenchido com as coordenadas geográficas. Observando a figura A.4, verifica-se que só 0.08% das ocorrências apresentam as coordenadas geográficas na descrição do local. Foi verificado o número de nulos existentes no código da via e apenas 0.08% das ocorrências não apresentam valores neste atributo. As ocorrências que apresentam as coordenadas do local são as ocorrências que não têm o código da via.

O atributo do código SIG representa o edifício/casa da ocorrência, mas caso a ocorrência não tenha acontecido dentro de um edifício/casa, este atributo pode apresentar, novamente, o código da via.



Figura A.5: Número de ocorrências com e sem código SIG.

A figura A.5 mostra o número de ocorrências, do Regimento de Sapadores Bombeiros, que apresentam, e que não apresentam, o código SIG do local da ocorrência. Analisando a figura A.5, verifica-se que 10.95% das ocorrências não têm valores neste atributo. Nunca acontece o atributo do código SIG estar preenchido e o código da via estar a nulo. De notar que, ocorrências que tenham ocorrido na rua, podem apresentar neste atributo, novamente, o código da via.

O código SIG da fracção representa a fracção do edifício do local da ocorrência, por exemplo, o andar.

A figura A.6 apresenta a distribuição dos valores do código de fracção do local das ocorrências do Regimento de Sapadores bombeiros. A figura A.6 mostra que 81.25% das ocorrências não apresentam código da fracção, 16.65% apresentam o código SIG da fracção e 2,10% apresentam a descrição da fracção, isto é, apresentam, por exemplo, o andar por escrito e não em código. Isto revela inconsistência com o apresentado nos



Figura A.6: Distribuição dos valores de código de fracção.

metadados. Existe um outro atributo, o atributo de complemento do local, que em muitos dos casos, apresenta a informação da fracção por escrito.

Com a exploração destes atributos foram detectados vários problemas. Todos estes atributos apresentam nulos, no entanto, o que apresenta maior percentagem é o código SIG da fracção. Foram detectados valores do código da via que devem ser posteriormente abordados. Existem problemas no atributo da fracção, pois segundo os metadados os valores deste atributo são só códigos, no entanto, isto não se verifica.

### A.3.5 Freguesia

Existem dois atributos referentes à freguesia da ocorrência, o identificador da freguesia e a descrição da mesma.



Figura A.7: Número de ocorrências que apresentam e não apresentam identificadores da freguesia.



Figura A.8: Número de ocorrências que apresentam e não apresentam designação da freguesia.

As figuras A.7 e A.8 mostram o número de ocorrências, do Regimento de Sapadores Bombeiros, que apresentam, e que não apresentam, informações sobre o identificador ou a designação da freguesia do local da ocorrência, respectivamente. Como a figura A.7 demonstra 6.73% das ocorrências não apresentam identificador da freguesia. No entanto, ao analisar a figura A.8 apenas 0.18% das ocorrências não apresentam a designação da freguesia. Estas informações são um alerta, pois o identificador deveria mapear para a



designação da freguesia, no entanto, existe um maior número de nulos no identificador e menos nulos na designação da freguesia. Isto revela que, em alguns casos, a designação da freguesia se encontra disponível mesmo sem identificador.

Ao explorar os dados referentes aos identificadores das freguesias verificou-se, ainda, que não são únicos, isto é, existe mais do que um identificador a identificar a mesma descrição de freguesia. A razão por detrás destes facto foi a reorganização administrativa de 2012 que levou à união de várias freguesias e, consequentemente, a uma nova identificação das freguesias no sistema.

Com a exploração destes dois atributos foram detectados problemas nos dados. Diferentes identificadores identificam a mesma designação de freguesia. Existem ocorrências que apresentam descrição de freguesia mas não têm nenhum identificador de freguesia associado.

### A.3.6 Estado da ocorrência

Existem três atributos referentes ao estado da ocorrência, o identificador, a descrição e a data e hora de alteração do estado da ocorrência.

Ao explorar os dados referentes aos identificadores do estado das ocorrências verificou-se que cada identificador mapeia para uma e uma só descrição do estado da ocorrência.



Figura A.9: Distribuição da descrição do estado da ocorrência.

A figura A.9 apresenta a distribuição dos valores da descrição do estado das ocorrências do Regimento de Sapadores Bombeiros. Como é observável na figura A.9, a maioria das ocorrências, mais de 98%, apresentam o estado de fecho administrativo, isto é, a ocorrência já esteve accionada, já foi fechada operacionalmente e já foi produzido um relatório relativo à intervenção. Cerca 2% das ocorrências estão em fecho operacional, isto é, a intervenção no terreno acabou. Apenas seis ocorrências ainda estão accionadas, isto é, a intervenção está em curso. No entanto, visto que o ano actual é 2020, é estranho que as intervenções ainda estejam accionadas. Investigando mais, as ocorrências que se encontram no estado de accionadas são todas referentes ao dia 17 de Dezembro de 2018, ou seja, o último dia dos dados. As alterações do estado destas ocorrências só devem ser visíveis nos dados do ano seguinte, de 2019.

Para validação dos dados, foi analisado se a data de alteração do estado é sempre posterior à data de início da ocorrência e verificou-se que é sempre posterior.

Com a exploração destes atributos não foram detectados problemas com os dados. No entanto, por uma questão de aproximação a realidade, pode-se ponderar mudar o estado

das ocorrências accionadas para fechadas operacionalmente, visto já ter passado mais de um ano e a intervenção no campo já ter, possivelmente, terminado.

### A.3.7 Tipologia da ocorrência

Existem dois atributos referentes à tipologia da ocorrência, o identificador e a descrição da tipologia da ocorrência.

O identificador que representa a tipologia da ocorrência não apresenta valores a nulo, ou seja, existe em todas as ocorrências um identificador do tipo.

Foram detectadas três situações problemáticas relativas aos identificadores. O identificador de tipologia 176 tem duas descrições da tipologia diferentes associadas, "8711 - Actividades - Apoio Social Avaliação" e "8711 - Actividades - Apoio Social Visitas Pontuais". O mesmo acontece com o identificador 177, tem duas descrições da tipologia diferentes associadas. O outro identificador problemático é o 42, pois não tem nenhuma descrição da tipologia associada.

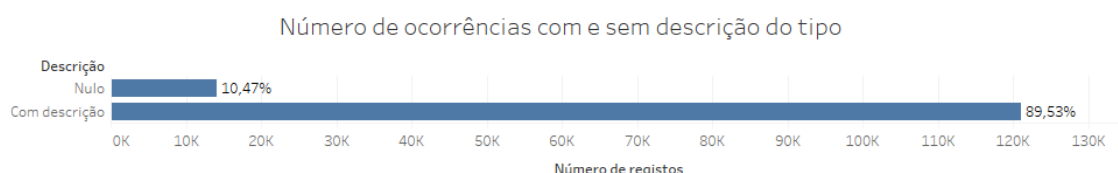


Figura A.10: Número de ocorrências que apresentam e não apresentam descrição do tipo de ocorrência.

A figura A.10 apresenta o número de ocorrências, do Regimento de Sapadores Bombeiros, que apresentam, e que não apresentam, a descrição da tipologia da ocorrência. Ao analisar a figura A.10 verifica-se que 10.47% das ocorrências não apresentam descrição da tipologia, no entanto apresentam identificador da tipologia. Para investigar mais esta questão foram analisados os anos onde este fenómeno aconteceu, os resultados estão presentes na figura A.11. Ao analisar a figura A.11 verifica-se que apenas os dados referentes aos anos de 2011 e 2012 sofrem deste problema. Indo ainda mais ao pormenor, é possível encontrar ocorrências, relativas aos anos com este problema e com identificadores de tipologia iguais, com descrição e outras sem descrição.

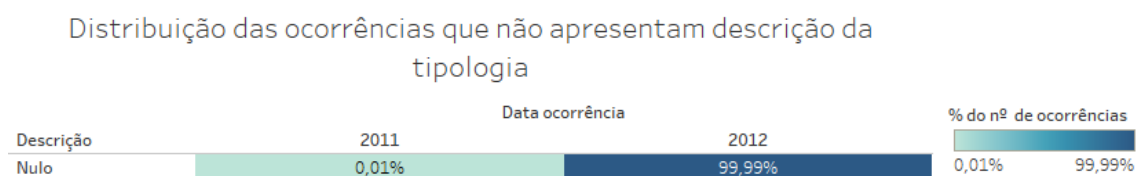


Figura A.11: Distribuição das ocorrências que não apresentam descrição da sua tipologia ao longo dos anos.

Ainda sobre a descrição da tipologia, ao analisar as várias tipologias existentes, foi detectada uma hierarquia. É possível caracterizar uma ocorrência com uma tipologia

mais geral. Consoante essa tipologia mais geral existem vários subtipos associados. A tabela B.1 apresenta quais são os subtipos existentes para ocorrências com a tipologia geral "Pré-Hospitalar".

Tabela A.3: Subtipos das ocorrências da tipologia geral "Pré-Hospitalar"

Tipo	Subtipo
Pré-Hospitalar	Afogamento
	Doença Súbita
	Intoxicação
	Parto
	Queimado
	Traumatismo/Queda

Com a exploração destes atributos foram detectados alguns problemas. Existem dois identificadores com diferentes descrições da tipologia. Acontece que algumas ocorrências, todos nos anos de 2012 e 2011, o atributo do identificador da tipologia estar preenchido, mas o atributo da descrição dessa tipologia estar em branco. Há um identificador que nunca tem uma descrição da tipologia associada. Foi também detectada uma hierarquia nas tipologias, é possível agrupar as ocorrências por um tipo mais geral e depois separar em subtipos mais específicos.

### A.3.8 Entidades

Existem dois atributos referentes à entidade que iniciou a ocorrência, o identificador e a descrição da entidade.

Os identificadores que identificam as entidades que deram início à ocorrência são únicos, isto é, cada identificador corresponde a uma e uma só descrição da entidade.

A figura A.12 representa a distribuição da descrição da entidade que iniciou a ocorrência do Regimento de Sapadores Bombeiros. Ao analisar a figura A.12 verifica-se que mais de metade das ocorrências são relatadas pelo público, no entanto, também se verifica que cerca de 3% das ocorrências não apresentam entidade associada. De notar que existe uma entidade denominada "Nenhuma". A existência desta entidade levanta dúvidas acerca do significado dos valores a nulo.

Ao explorar este atributo verificou-se que este apresenta problemas de completude dos dados. Existem algumas ocorrências que não apresentam a entidade que iniciou o processo da ocorrência. No entanto, a existência de uma entidade "Nenhuma" levanta algumas questões sobre o significado dos valores a nulo.

### A.3.9 Indicador do grau de risco

Este atributo indica o grau de risco de uma ocorrência. Segundo os metadados disponibilizados, o valor de 1 neste atributo revela uma ocorrência com um grau de risco elevado, enquanto o valor de 2 representa uma ocorrência com risco moderado.

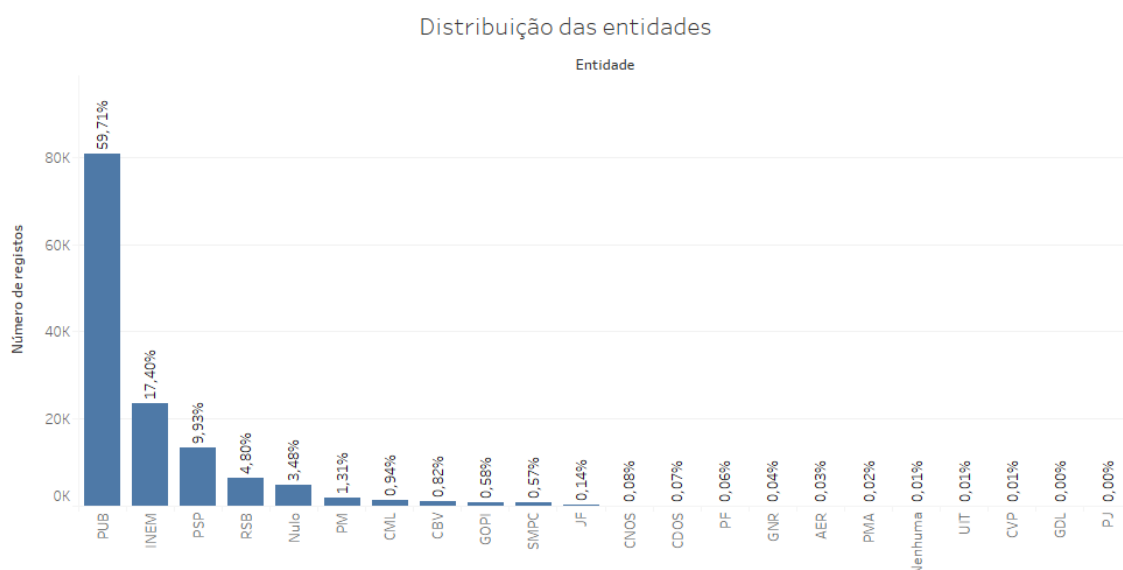


Figura A.12: Distribuição dos valores da entidade que deu início à ocorrência.

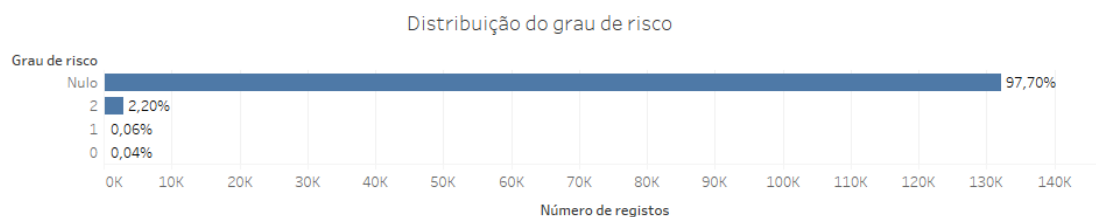


Figura A.13: Distribuição dos valores do grau de risco da ocorrência.

A figura A.13 representa a distribuição dos valores do grau de risco das ocorrências do Regimento de Sapadores Bombeiros. Analisando a figura A.13, verifica-se que existe mais um valor que não está presente nos metadados, o valor 0. É possível também observar que este atributo é maioritariamente nulo, mais de 97% das ocorrências não têm grau de risco associado.

Com a exploração deste atributo conclui-se que falta informação nos metadados, não é especificado o que significa uma ocorrência ter grau de risco 0. Devido à percentagem alta de valores nulos, a utilidade deste atributo para futura análise dos dados é reduzida.

### A.3.10 Indicador de suspensão antes da activação de meios

Este atributo indica se a ocorrência foi, ou não, suspensa antes da activação de meios para o local. Sendo este atributo um booleano, apenas existem dois valores possíveis.

A figura A.14 representa a distribuição dos valores do indicador de suspensão antes da activação de meios para o local da ocorrência do Regimento de Bombeiros Sapadores. Analisando a figura A.14, verifica-se que todas as ocorrências contêm este atributo preenchido. Mais de 97% das ocorrências registadas não foram suspensas antes da activação e apenas 2.86% foram suspensas.



Figura A.14: Distribuição dos valores do indicador de suspensão antes da activação de meios.

Com a exploração deste atributo não foi detectado nenhum problema nos dados.

### A.3.11 Indicador de falso alarme depois da activação de meios

Este atributo indica se a ocorrência foi falso alarme depois da activação de meios para o local. Sendo este atributo um booleano, apenas existem dois valores possíveis.

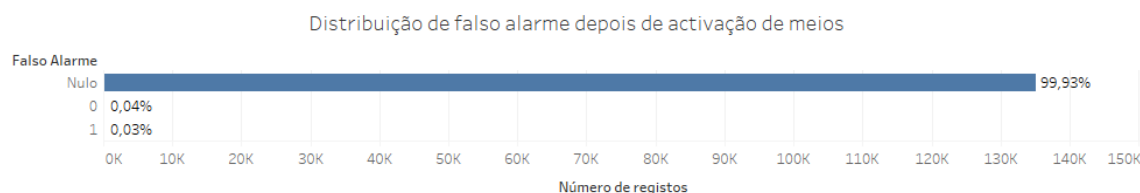


Figura A.15: Distribuição dos valores de falso alarme depois da activação de meios.

A figura A.15 representa a distribuição dos valores de falso alarme após a activação de meios para o local da ocorrência do Regimento de Sapadores Bombeiros. Observando a figura A.15 verifica-se que a maioria das ocorrências registadas, 99.93% das ocorrências, não apresentam informação neste atributo.

Com a exploração deste atributo conclui-se que o mesmo não deve ser usado para análise dos dados devido à elevada percentagem de valores a nulo.

### A.3.12 Número de viaturas

Este atributo apresenta o número de viaturas que serão enviadas para o local da ocorrência.

A figura A.16 apresenta a distribuição dos valores do número de viaturas enviadas para o local da ocorrência do Regimento de Sapadores Bombeiros. Analisando a figura A.16 observa-se que na maioria das ocorrências, cerca de 85%, apenas é enviada uma viatura para o local. Existem ocorrências para as quais foram mobilizadas muitas viaturas, no entanto, essas ocorrências são poucas. De notar que existem ocorrências que não têm viaturas a ser enviadas para o local.

A figura A.17 mostra a relação que existe entre a suspensão da ocorrência e o não envio de viaturas para o local da ocorrência. Analisando a figura A.17 observa-se que existe uma relação entre o não envio de viaturas e a suspensão da ocorrência. Em 97.32% das ocorrências, que não tiveram viaturas enviadas para o local, a causa do não envio de

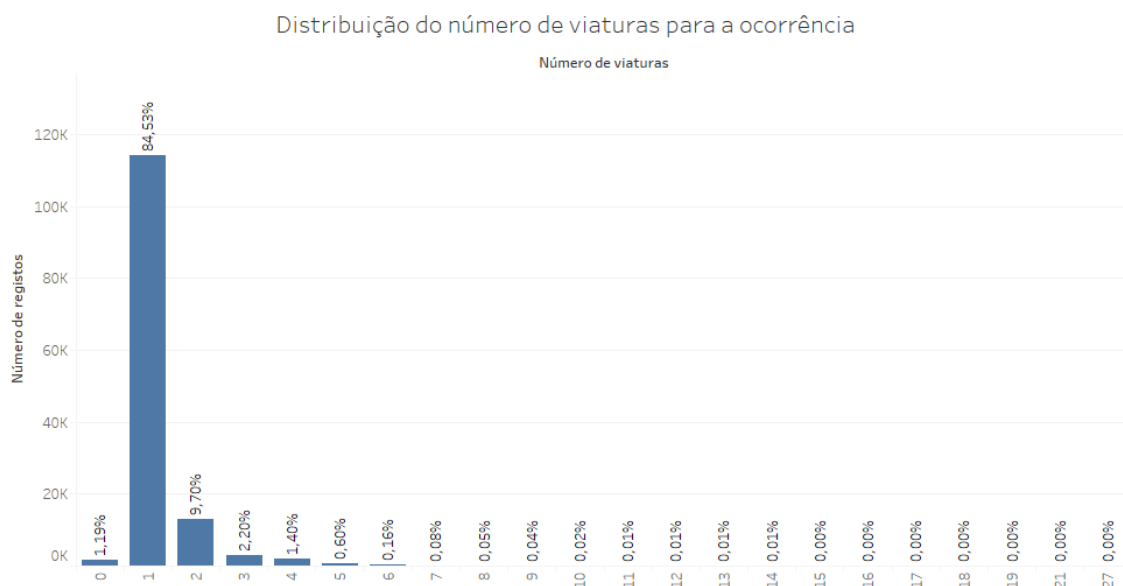


Figura A.16: Distribuição dos valores do número de viaturas enviadas para o local.

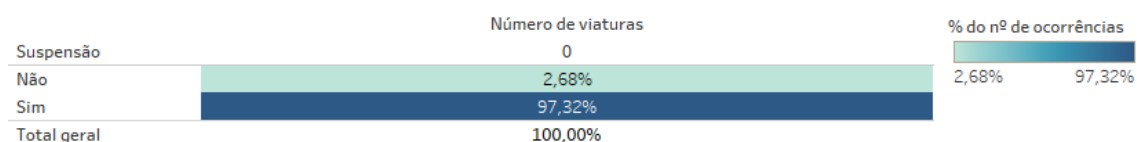


Figura A.17: Relação entre nenhuma viatura enviada para o local e a suspensão da ocorrência.

viaturas foi a suspensão da ocorrência antes da activação de meios. No entanto, existem 2.68% de ocorrências não suspensas que não tiveram viaturas a ser enviadas para o local.

Com a exploração deste atributo conclui-se que existe uma relação entre o atributo de suspensão da ocorrência e o número de viaturas enviadas para o local, no entanto, ainda existem ocorrências que mesmo não sendo suspensas não têm meios mobilizados. Como existem ocorrências onde são dispensadas várias viaturas, seria interessante relacionar essas ocorrências com o seu tipo geral. Este tópico será abordado na seguinte subsecção.

### A.3.13 Número estimado de elementos

Este atributo apresenta os valores do número estimado de elementos enviados para o local da ocorrência.

A figura A.18 apresenta a distribuição dos valores da estimativa do número de elementos a enviar para o local da ocorrência do Regimento de Sapadores Bombeiros. Ao analisar a figura A.18 verifica-se que o número mais usual da estimativa de elementos é o valor seis, com 37.84% das ocorrências. Existem ocorrências para as quais foi estimado um número muito elevado de elementos, no entanto, essas ocorrências são poucas. Existem cerca de 17% de ocorrências que não apresentam estimativa do número de elementos a enviar para o local. Este facto poderia estar relacionado com a suspensão da ocorrências,

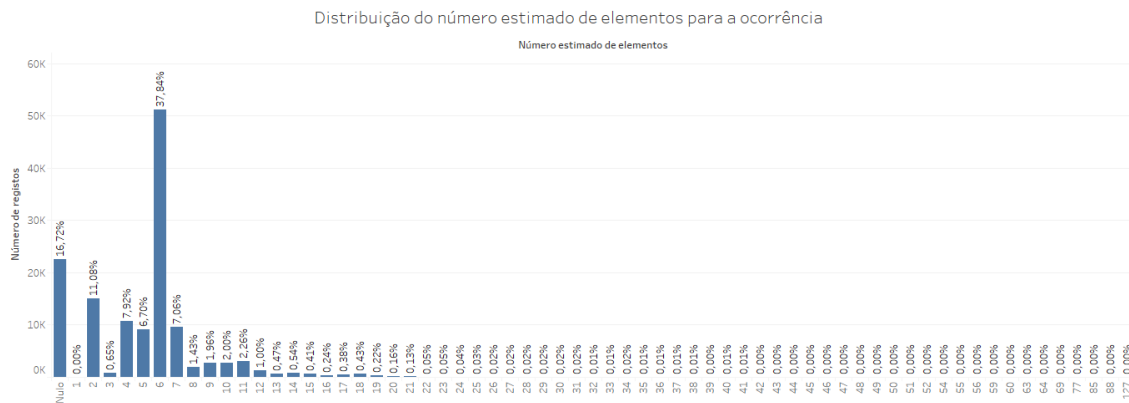


Figura A.18: Distribuição dos valores da estimativa do número de elementos enviados para o local.

mas, após análise, não se verifica nenhuma relação.

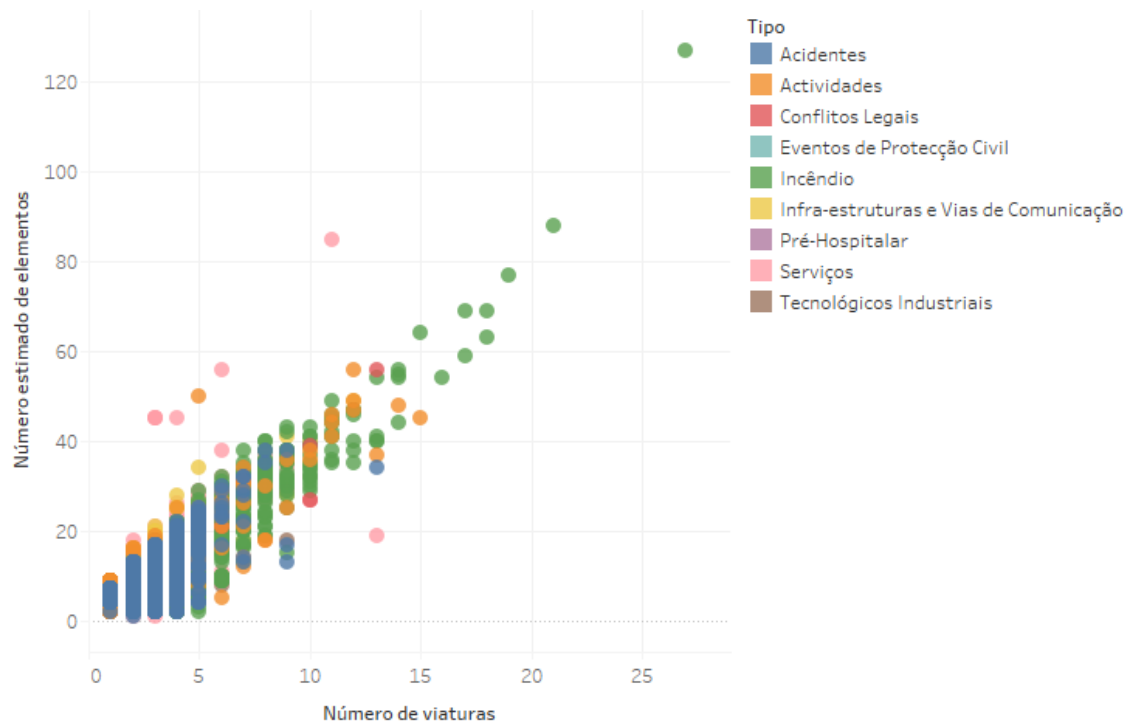


Figura A.19: Relação entre os atributos número estimado de elementos, número de viaturas e tipologia.

A figura A.19 apresenta a relação entre os atributos número estimado de elementos, número de viaturas a enviar para o local e a tipologia da ocorrência do Regimento de Sapadores Bombeiros. Ao analisar a figura A.19 é perceptível uma relação linear entre o número de viaturas e número estimado de elementos. A tipologia geral tem também uma relação com estes atributos, nomeadamente, as ocorrências relativas a incêndios são as que apresentam maior número de viaturas e estimativa de elementos mobilizados para o local.

Com a exploração deste atributo verifica-se a presença de alguns valores nulos. Este atributo apresenta uma relação tanto com o atributo do número de viaturas a ser enviadas para o local, como com a tipologia da ocorrência.

#### **A.3.14 Descrição da ocorrência**

O atributo da descrição da ocorrência é um campo de texto livre onde se tem uma descrição mais detalhada da ocorrência. Este atributo apresenta algumas informações que podem ser relevantes para análise futura. Algumas dessas informações, no caso de ocorrências médicas, são a idade do paciente e códigos [CODU](#).

### **A.4 Verificação da qualidade dos dados**

Verificou-se que estes dados apresentam alguns problemas de qualidade. A maior parte destes problemas estão relacionados com a completude dos dados, mas também existem problemas de consistência, exactidão e validação. Não foram detectado problemas de duplicados.

#### **A.4.1 Completude**

Os atributos que apresentam problemas de completude são:

- Latitude e longitude;
- Código da Via;
- Código SIG;
- Código SIG da fracção;
- Identificador e descrição da freguesia;
- Identificador e descrição da tipologia;
- Identificador e descrição da entidade;
- Indicador grau de risco;
- Indicador de falso alarme depois da activação de meios;
- Estimativa do número de elementos.

Relativamente à latitude e longitude, existem ocorrências que não apresentam coordenadas geográficas. Para evitar a existência de nulos na latitude e longitude e, conseqüentemente, aumentar a qualidade destes dados, poderão ser geradas coordenadas aproximadas para as ocorrências que apresentem morada associada.



Em relação ao código da via, este atributo apresenta alguns valores a nulo e sempre que isso acontece a descrição do local apresenta coordenadas geográficas. Como todas estas ocorrências apresentam coordenadas geográficas, a ausência deste código não é um problema grave, pois continua-se a ter informação sobre onde a ocorrência aconteceu.

Relativamente ao código SIG, este apresenta também alguns valores a nulo. Sendo que este atributo pode, ou não, conter o código de via quando a ocorrência acontece na rua, será que alguns dos valores a nulo representam que a ocorrência aconteceu na rua? Ou que, simplesmente, não se sabe o edifício? Estas questões têm de ser ponderadas e os metadados poderiam ser mais explícitos, pois apenas diz que pode, ou não, conter códigos de via.

Quanto ao código SIG da fracção, este apresenta uma percentagem alta de valores a nulo. Este atributo pode estar a nulo, em alguns casos, devido à ocorrência ter acontecido na rua. Para aumentar a qualidade dos dados, a utilização deste atributo para analisar os dados no futuro deve ser ponderada.

Relativamente às freguesias, tanto o identificador como a descrição da freguesia apresentam valores nulos. No entanto, os identificadores apresentam uma percentagem maior de nulos comparado com a descrição. Foram detectadas ocorrências onde a designação da freguesia está presente, mas o identificador não. Para melhorar a qualidade destes dados, deve-se ponderar preencher os identificadores com os valores que dizem respeito à designação da freguesia presente na ocorrência. Para ocorrências que não apresentem nem o identificador nem a descrição da freguesia, mas apresentem morada é possível associar a morada a uma freguesia.

Relativamente aos problemas da tipologia, existe um identificador que não apresenta a descrição da tipologia da ocorrência. Para melhorar o nível de qualidade, é possível que ocorrências com este identificador de tipologia tenham de ser descartadas. Outro problema recorrente, principalmente nas ocorrências referentes ao ano de 2012, é algumas ocorrências terem o atributo do identificador preenchido mas depois não terem a descrição da tipologia. Para melhorar este aspecto, terá de ser tomada a decisão de preencher as descrições assumindo que os identificadores referem-se sempre à mesma descrição.

Quanto à entidade que iniciou o processo da ocorrência, tanto o identificador como a descrição da entidade apresentam valores nulos. A existência de uma entidade denominada "Nenhuma" levanta questões do significado dos valores a nulo. Será que os valores a nulo são considerados o mesmo que a entidade "Nenhuma"? Esta questão tem de ser apurada, só depois de apurada poderá ser tomada uma decisão, nomeadamente, de se juntar as ocorrências sem entidade à entidade "Nenhuma".

Relativamente ao indicador grau de risco, este atributo apresenta uma percentagem muito alta de valores a nulo. Para melhorar a qualidade destes dados, deve-se ponderar a não utilização deste atributo numa futura análise.

Em relação ao indicador de falso alarme depois da activação de meios para o local da ocorrência, este atributo apresenta uma percentagem muito alta de nulos. Este indicador é um booleano. Será que os valores a nulo são o mesmo que dizer que a ocorrência não foi

um falso alarme? Para melhorar a qualidade destes dados, deve-se ponderar a utilização deste atributo numa futura análise e procurar uma resposta plausível para a questão junto dos responsáveis dos dados.

Por fim, o atributo relativo à estimativa do número de elementos enviados para o local da ocorrência apresenta alguns valores a nulo. Para melhorar a qualidade destes dados, terá de ser tomada uma decisão sobre o que fazer com os valores nulos, se se mantêm ou se preenchem com alguma técnica de substituição, por exemplo, preencher com a média.

### A.4.2 Consistência

Os atributos que apresentam problemas de consistência são:

- Latitude e longitude;
- Identificadores da freguesia;
- Identificadores da tipologia.

Encontram-se ocorrências que não apresentam coordenadas geográficas. No entanto, estas apresentam outros atributos de onde é possível extrair a informação relativa à localização. Isto releva inconsistência e incompletude nos dados. Para evitar a existência de nulos nas coordenadas geográficas das ocorrências e, consequentemente, aumentar a qualidade destes dados, poderão ser geradas coordenadas aproximadas para as ocorrências que apresentem morada.

Existem problemas com os identificadores e designação das freguesias. Identificadores diferentes referem-se há mesma freguesia, revelando inconsistência nos dados. Isto acontece-se, porque, devido à união das freguesias, houve necessidade de fazer uma nova identificação das mesmas no sistema. Para aumentar a qualidade destes dados, deve-se adoptar ou o antigo, ou o novo, modelo de caracterização das freguesias de modo a atingir a consistência nos dados.

Existem dois identificadores da tipologia que se referem a diferentes descrições. Isto demonstra inconsistência. Para melhorar a qualidade destes dados, terá de ser tomada uma decisão. Uma possível solução passa por escolher uma das diferentes descrições para cada um desses identificadores ou apenas manter as palavras em comum das diferentes descrições.

### A.4.3 Validação

Os atributos que apresentam problemas de validação são:

- Latitude e longitude;
- Designação do local;
- Código SIG da fracção;

- Indicador do grau de risco.

Existem duas ocorrências que apresentam as coordenadas geográficas erradas. As coordenadas erradas referem-se ao Mar Mediterrâneo e à Holanda. Estas ocorrências devem ser descartadas numa futura análise.

Estão presentes valores das coordenadas geográficas na designação do local da ocorrência. Este atributo deveria-se referir, somente, à descrição do local da ocorrência, isto é, a morada.

Relativamente ao código SIG da fracção da ocorrência, existe conflito com o que está presente nos metadados. Nos metadados este atributo é só um código SIG. No entanto, algumas ocorrências apresentam valores simbólicos referindo-se, por exemplo, ao 2º piso do edifício. Existe um outro atributo, o complemento do local, que apresenta por escrito a fracção da ocorrência.

Em relação ao indicador do grau de risco, os metadados não espelham o que está presente nos dados, existe falta de informação. Os metadados referem a existência de apenas 2 valores possíveis, um para o risco elevado e outro para o risco moderado. No entanto, nos dados existem 3 valores possíveis para este atributo. Os metadados devem ser actualizados e atribuído um significado para o outro valor.

O atributo da data de alteração do estado da ocorrência foi analisado para validação. A data da ocorrência tem sempre de ser anterior à data de alteração do estado. Foi verificado e nenhuma data de alteração do estado é anterior à data da ocorrência, não havendo assim problemas.

##### A.4.4 Precisão

Problemas de precisão detectados em:

- Estado da ocorrência;
- Ocorrências do ano de 2011.

Quanto ao estado da ocorrência, não existem problemas de consistência nem incompletude destes dados. No entanto, para aproximar os dados à realidade, é preciso olhar para as ocorrências de 2018 que estão accionadas e ponderar mudar o seu estado para fechadas operacionalmente.

Por último, existem problemas nos dados referentes ao ano de 2011. Este ano apresenta poucas ocorrências e existem meses onde não existem registos. Para aumentar a qualidade dos dados, este ano deve ser excluído, porque não é um ano representativo da realidade e não tem interesse para futuras análises.

#### **A.4.5 Outras notas**

Há que ter em conta que, relativamente às moradas, existem alguns valores do código da via que têm significados especiais. Os valores em questão são o 0 e valores começados por 9999999.

A tipologia das ocorrências apresentam hierarquias. É possível as ocorrências apresentarem uma tipologia geral e um subtipo mais específico associado à tipologia geral.

Existem ocorrências, que apesar de não terem sido suspensas, onde não são enviadas viaturas para o local. Estas ocorrências são suspeitas e necessitam de ser analisadas mais ao pormenor para decidir se são ocorrências a manter numa futura análise.

Por fim, o atributo da descrição da ocorrência é um campo de texto livre que apresenta informações que podem ser interessantes para uma futura análise. Será necessário desenvolver um método para extrair informação deste atributo.

## LISTAGEM DOS SUBTIPOS DAS TIPOLOGIAS GERAIS DAS OCORRÊNCIAS

A tabela B.1 apresenta quais são os subtipos associados às tipologias gerais das ocorrências do Regimento de Sapadores Bombeiros.

Tipologia Geral	Subtipos
Acidentes	<p>Aquático</p> <p>Aquático - Queda ao Rio</p> <p>Equipamentos</p> <p>Equipamentos - Elevadores</p> <p>Equipamentos - Escadas/Tap. Rolantes</p> <p>Ferroviário - Abalroamento</p> <p>Ferroviário - Atropelamento</p> <p>Ferroviário - Choque</p> <p>Ferroviário - Com Encarcerados</p> <p>Ferroviário - Descarrilamento</p> <p>Rodoviários - Atropelamento</p> <p>Rodoviários - C/ Encarcerados</p> <p>Rodoviários - Com viaturas</p>
Actividades	<p>Acompanhamento Consultas</p> <p>Acompanhamento Hospitalar</p> <p>Acompanhamento Retornos</p> <p>Apoio Social Avaliação</p> <p>Apoio Social Monitorização</p> <p>Apoio Social Sinalização</p> <p>Apoio Social Visitas Pontuais</p>

APÊNDICE B. LISTAGEM DOS SUBTIPOS DAS TIPOLOGIAS GERAIS DAS OCORRÊNCIAS

	<p>Apoio Social Visitas Regulares</p> <p>Assistência à População/Apoio Social</p> <p>Busca/Resgate (Pessoas e Animais) - Aquático</p> <p>Busca/Resgate (Pessoas e Animais) - Terrestre</p> <p>Busca/Resgate (Pessoas e Animais) Aquático</p> <p>Deslocações - Serviço Geral</p> <p>Evacuação</p> <p>Reuniões Eventos</p> <p>Tele assist. Documentação Expediente</p> <p>Tele assistência Adesão</p> <p>Tele assistência Manutenção Avarias</p> <p>Tele assistência Recolha Equipamentos</p>
Conflitos Legais	<p>Agressão/Violação</p> <p>Apoio à Autoridade</p> <p>Explosivos - Ameaça</p> <p>Explosivos - Explosão</p> <p>Suicídio/Homicídio - Consumado</p> <p>Suicídio/Homicídio - Tentativa</p> <p>Transporte/Remoção Cadáver</p>
Eventos de Protecção Civil	<p>Visita Técnica</p>
Incêndio	<p>Agrícola</p> <p>Detritos</p> <p>Edifício (Infra-estrutura/Instalação) - Comercial/Lojas/Feiras/Gare de Transporte</p> <p>Edifício (Infra-estrutura/Instalação) - Cultura/Museu/Arte/Biblioteca</p> <p>Edifício (Infra-estrutura/Instalação) - Edifício de grande altura (&gt;29 m)</p> <p>Edifício (Infra-estrutura/Instalação) - Edifício Devoluto/Degradado</p> <p>Edifício (Infra-estrutura/Instalação) - Escolar</p> <p>Edifício (Infra-estrutura/Instalação) - Espectáculo/Lazer/Culto Religioso</p> <p>Edifício (Infra-estrutura/Instalação) - Estacionamento</p> <p>Edifício (Infra-estrutura/Instalação) - Habitação</p> <p>Edifício (Infra-estrutura/Instalação) - Hospitalar/Lar</p> <p>Edifício (Infra-estrutura/Instalação) - Hoteleira e similar</p> <p>Edifício (Infra-estrutura/Instalação) - Indústria/Oficina/Armazém</p> <p>Edifício (Infra-estrutura/Instalação) - Militar/Forças Segurança</p> <p>Edifício (Infra-estrutura/Instalação) - Serviços</p> <p>Equipamentos (sem afectação do ambiente)</p> <p>Equipamentos (sem afectação do ambiente) - Contentores de lixo</p>

	<p>Inculto</p> <p>Povoamento Florestal</p> <p>Produtos</p> <p>Transportes - Aéreo</p> <p>Transportes - Ferroviário</p> <p>Transportes - Rodoviário</p> <p>Transportes Aquático</p>
Infra-estruturas e Vias de Comunicação	<p>Corte de abastecimento - Electricidade</p> <p>Danos/Queda Cabos Eléctricos</p> <p>Danos/Queda Cabos Eléctricos - Curto-circuito</p> <p>Desabamento</p> <p>Desabamento - Queda de Revestimento</p> <p>Desentupimento/Tamponamento</p> <p>Deslizamento</p> <p>Inundação</p> <p>Inundação Espaço Privado</p> <p>Inundação Espaço Público</p> <p>Queda de Árvore</p> <p>Queda de Estruturas</p>
Serviços	<p>Abastecimento de Água - Entidade Privada</p> <p>Abastecimento de Água - Entidade Pública</p> <p>Abastecimento de Água - População</p> <p>Abertura de Porta - Com Socorro</p> <p>Abertura de Porta - Sem Socorro</p> <p>Fecho de água</p> <p>Limpeza de Via/Conservação</p> <p>Limpeza de Via/Conservação - Óleo no Pavimento</p> <p>Limpeza de Via/Conservação - Sinalizar Buraco</p> <p>Prevenções - Desporto</p> <p>Prevenções - Espectáculo</p> <p>Prevenções - Patrulhamento/Vigilância</p> <p>Prevenções - Pré-Posicionamento Meios</p> <p>Prevenções - Queimadas</p> <p>Prevenções - Transportes</p> <p>Reboque/Desempanagem</p> <p>Resgate/Recolha de Animais</p> <p>Transporte Doentes - Auxílio p/ Transporte de Doentes</p> <p>Transporte Doentes - Geral</p>

APÊNDICE B. LISTAGEM DOS SUBTIPOS DAS TIPOLOGIAS GERAIS DAS OCORRÊNCIAS

	Transporte Doentes - Inter-Hospital
Tecnológicos Industriais	Acidentes Matérias Perigosas - Químicos  Fuga de Gás - Canalização/Conduta Fuga de Gás - Depósito/Reservatório Fuga de Gás - Garrafa Matérias Perigosas em Trânsito - Químicos Situações Suspeitas - Verificar Cheiros Situações Suspeitas - Verificar Fumos Situações Suspeitas - Verificar SADI/Alarmes

Tabela B.1: Subtipos das diversas tipologias gerais presentes no conjunto de dados das ocorrências do [RSB](#).





## LISTAGEM DOS DIAGRAMAS IVML DOS *DASHBOARDS* DESENVOLVIDOS

Este apêndice apresenta os diagramas [IVML](#) de interações e *tooltips* dos vários *dashboards* desenvolvidos nesta dissertação.

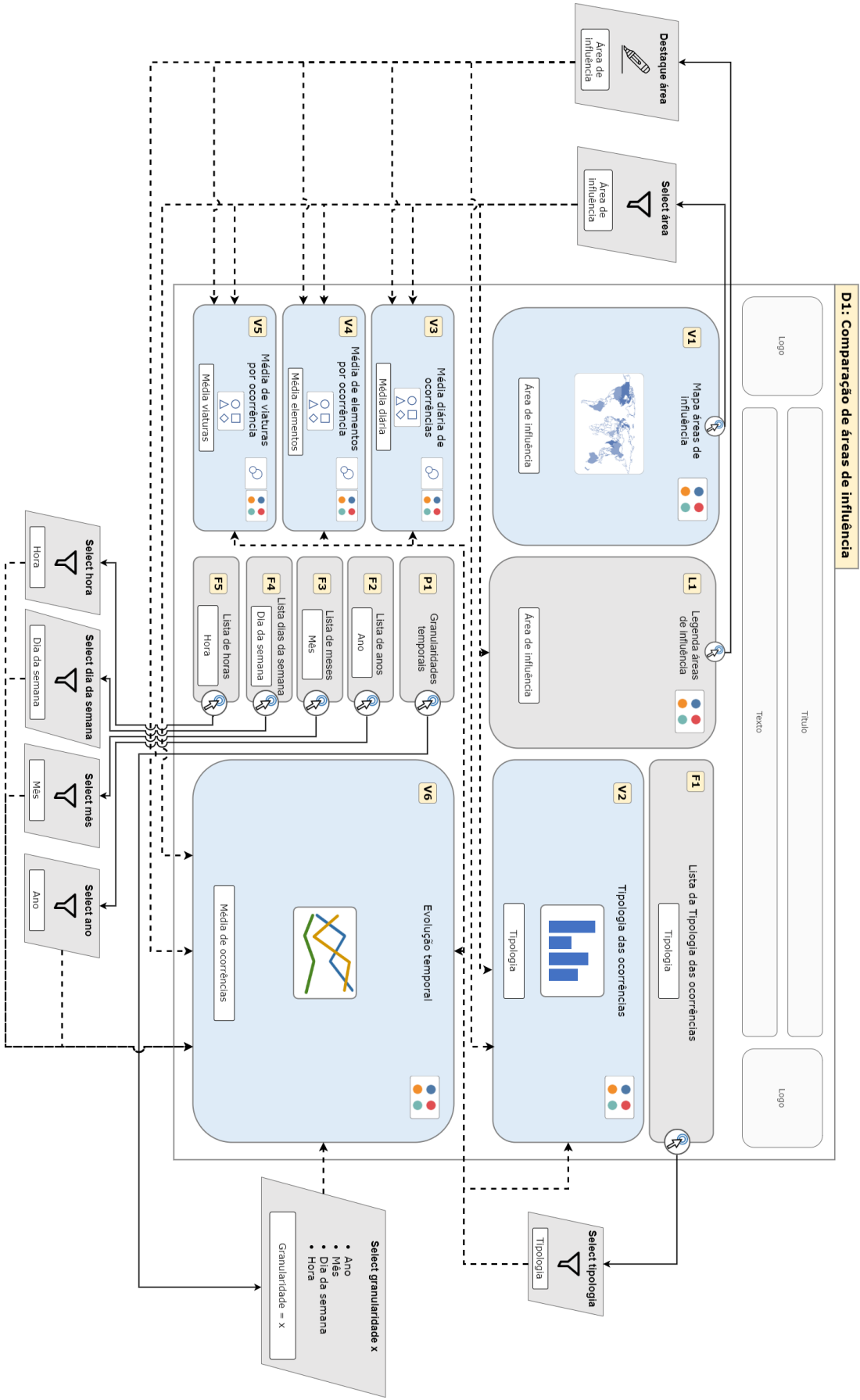


Figura C.1: Diagrama IVML de interações do *dashboard* de comparação das ocorrências nas áreas de influência dos quartéis do RSB.

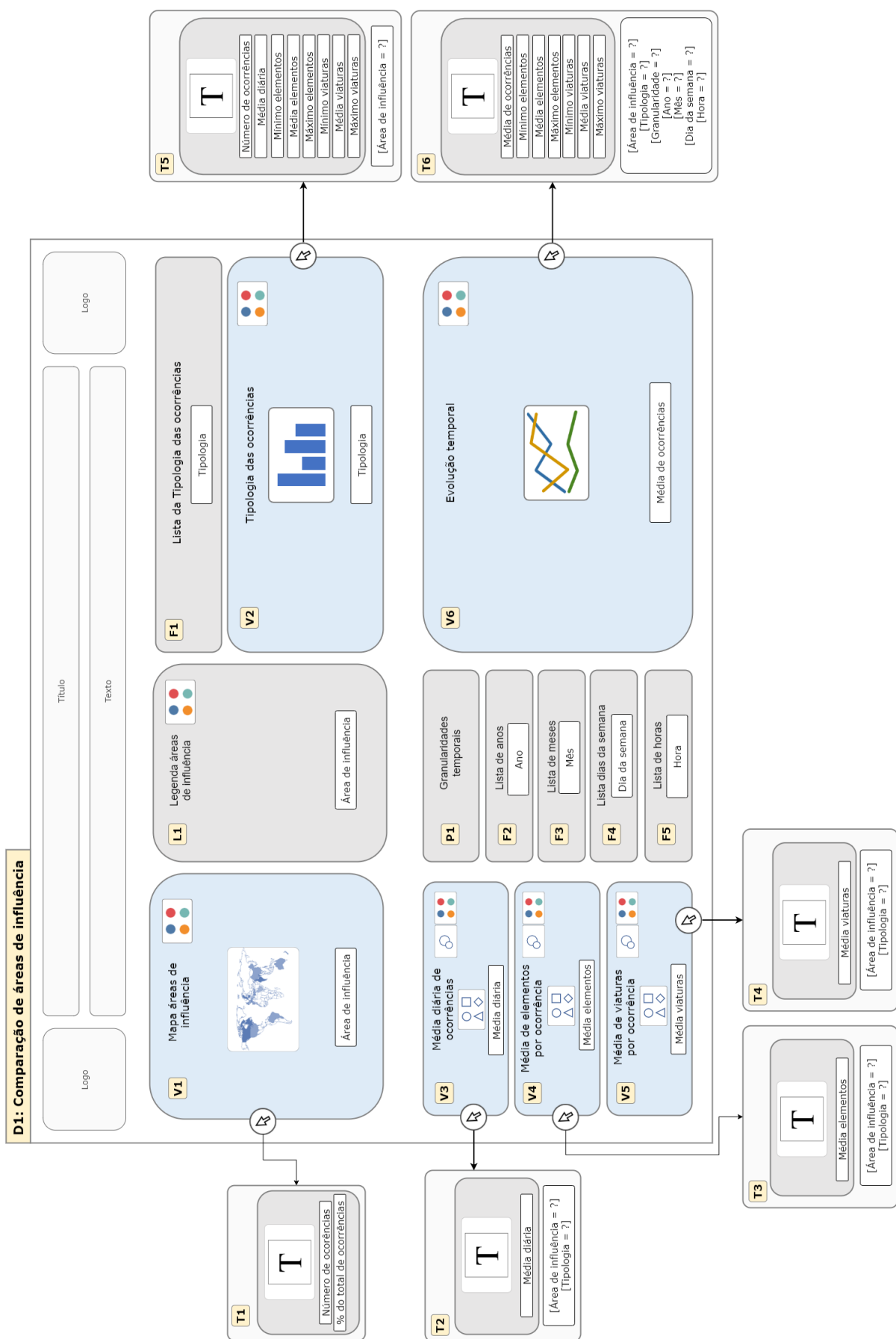


Figura C.2: Diagrama IVML de *tooltips* de comparação das ocorrências nas áreas de influência dos quartéis do RSB.

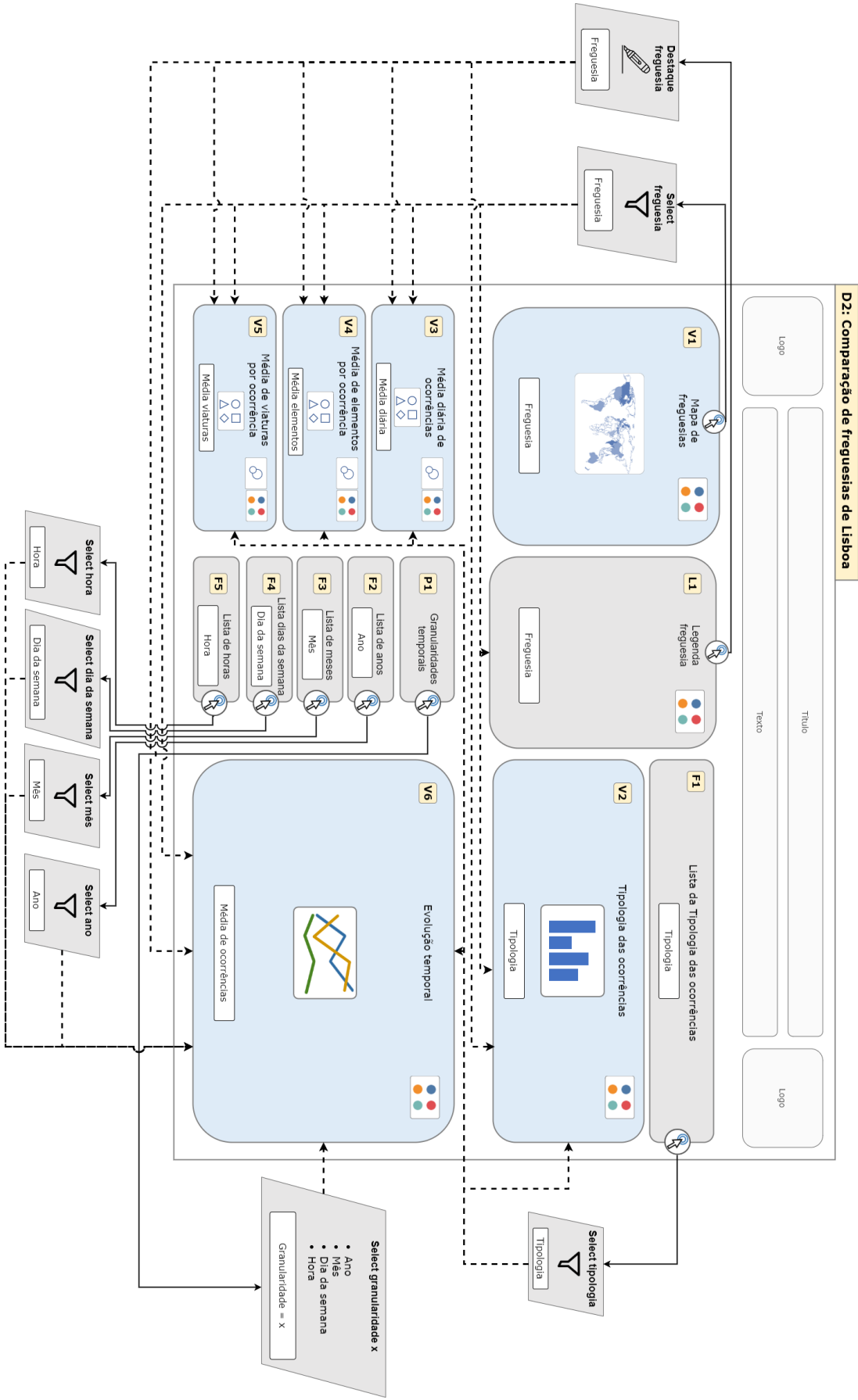


Figura C.3: Diagrama IVML de interações do *dashboard* de comparação das ocorrências nas freguesias do concelho de Lisboa.

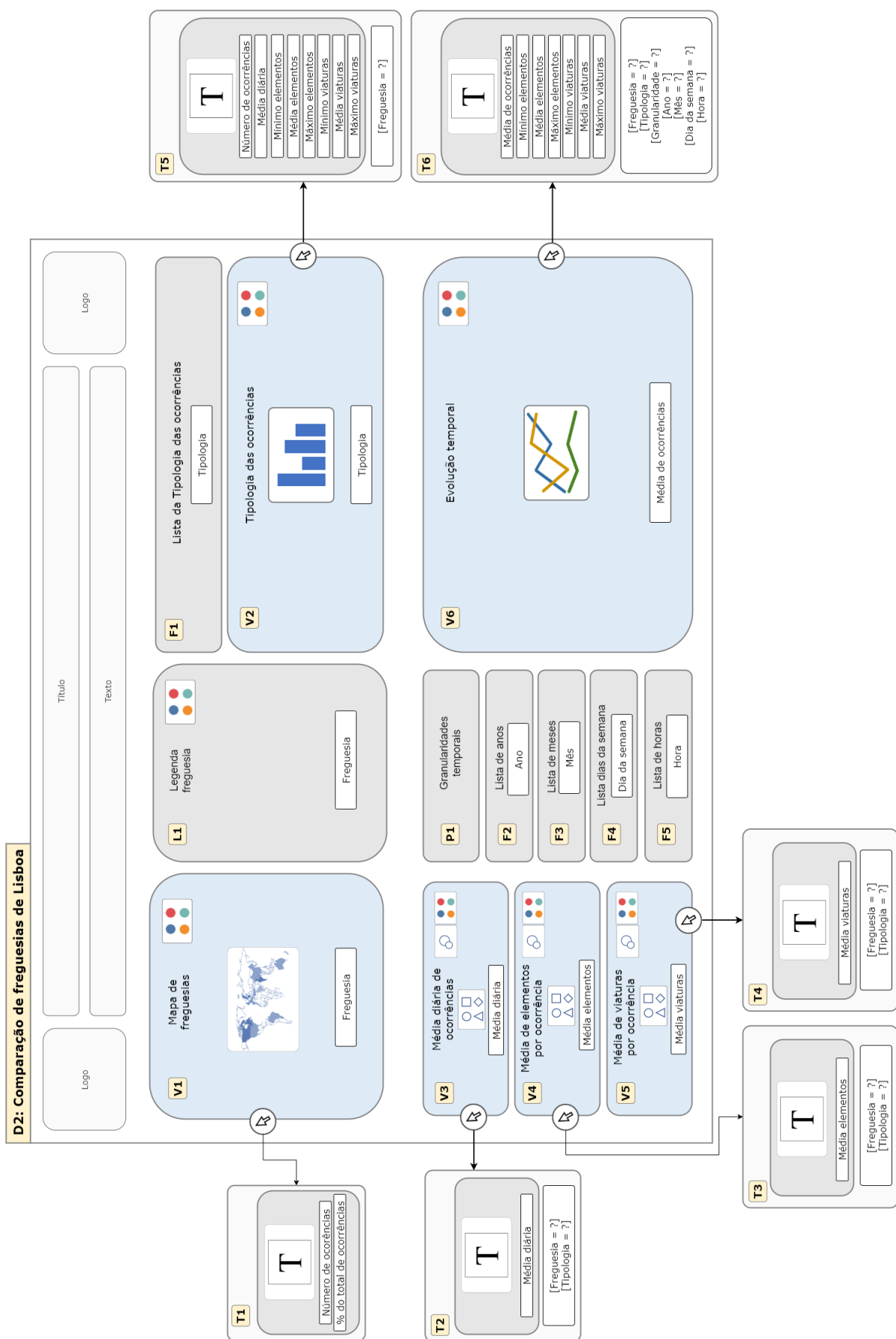


Figura C.4: Diagrama IVML de tooltips de comparação das ocorrências nas freguesias do concelho de Lisboa.

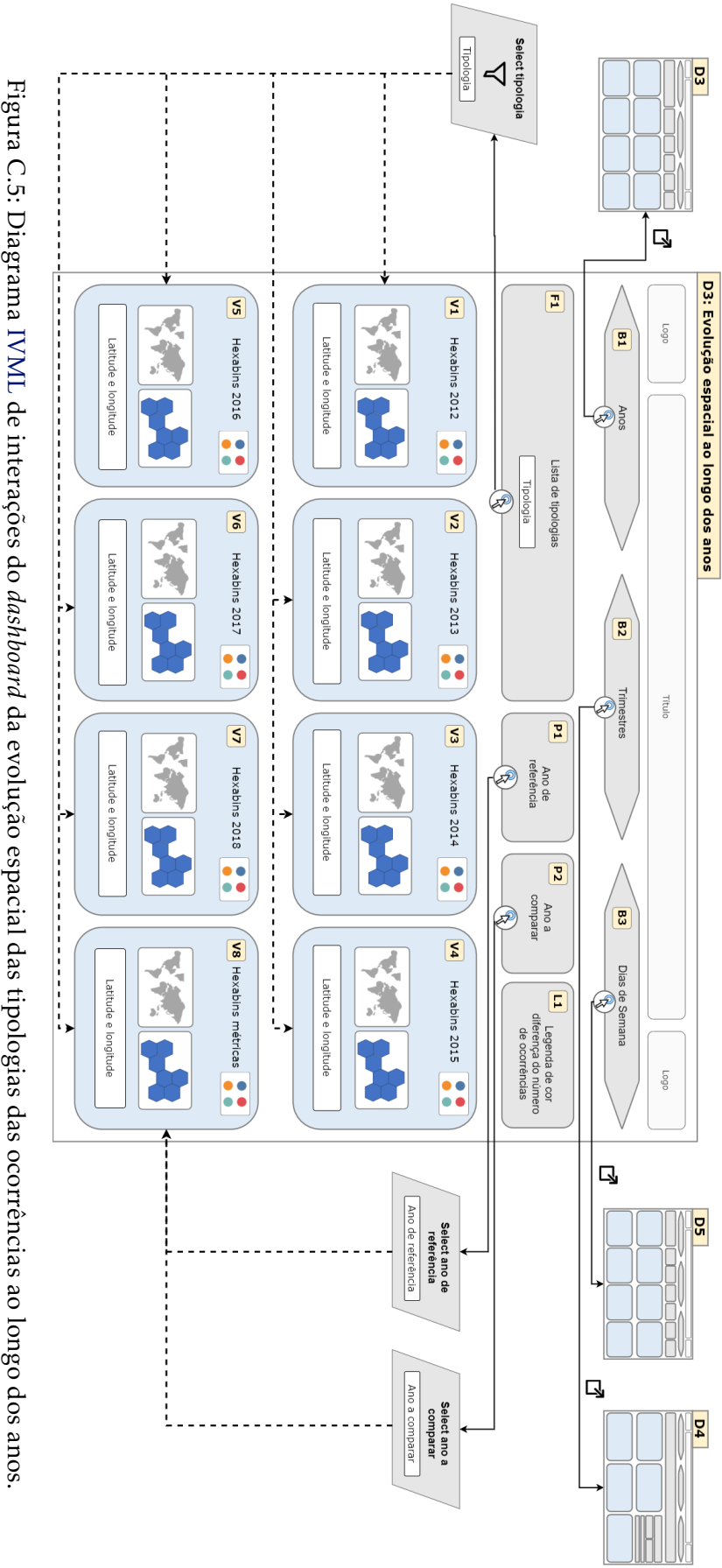


Figura C.5: Diagrama IVML de interações do *dashboard* da evolução espacial das tipologias das ocorrências ao longo dos anos.

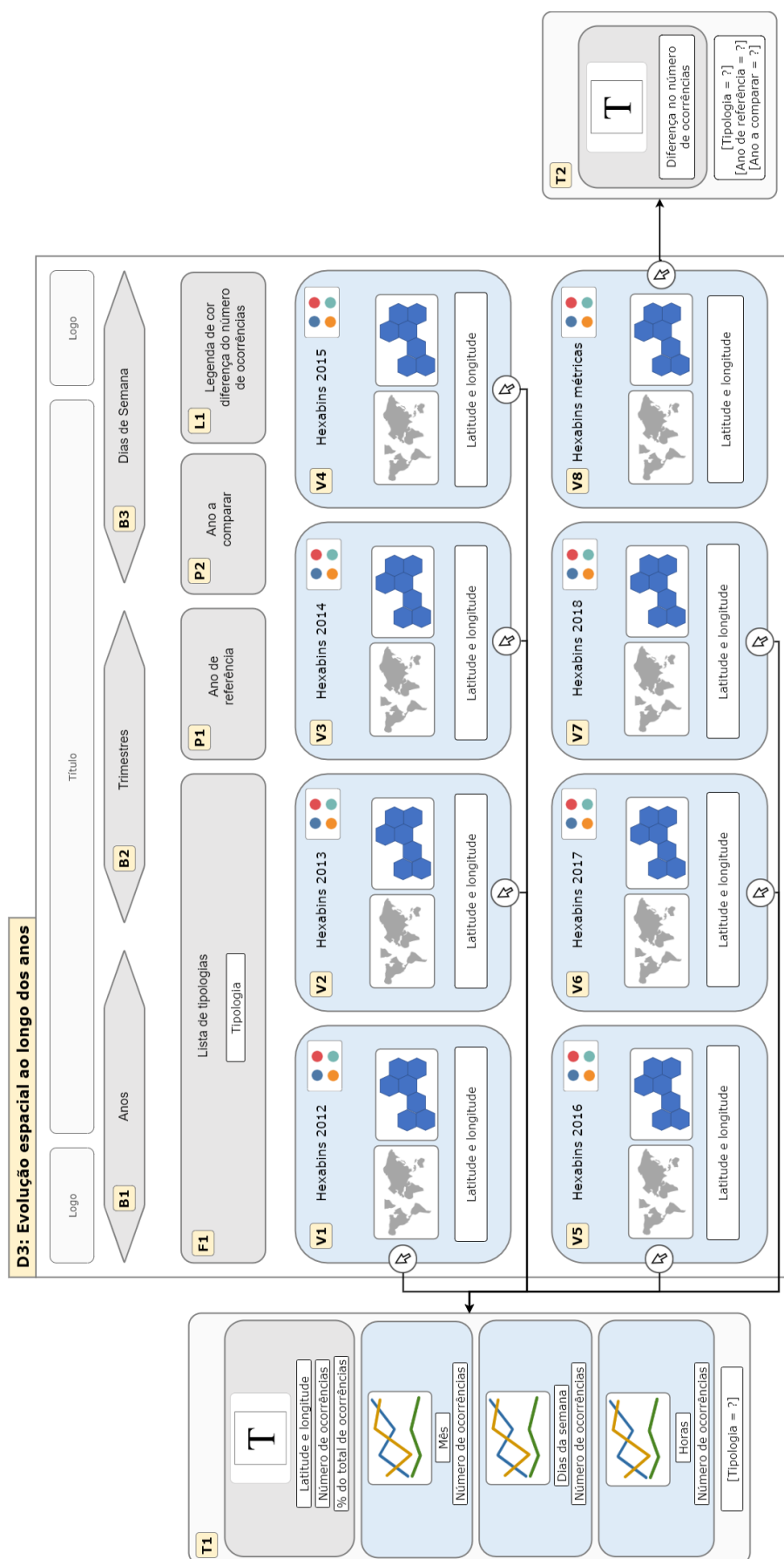


Figura C.6: Diagrama IVML de *tooltips* da evolução espacial das tipologias das ocorrências ao longo dos anos.

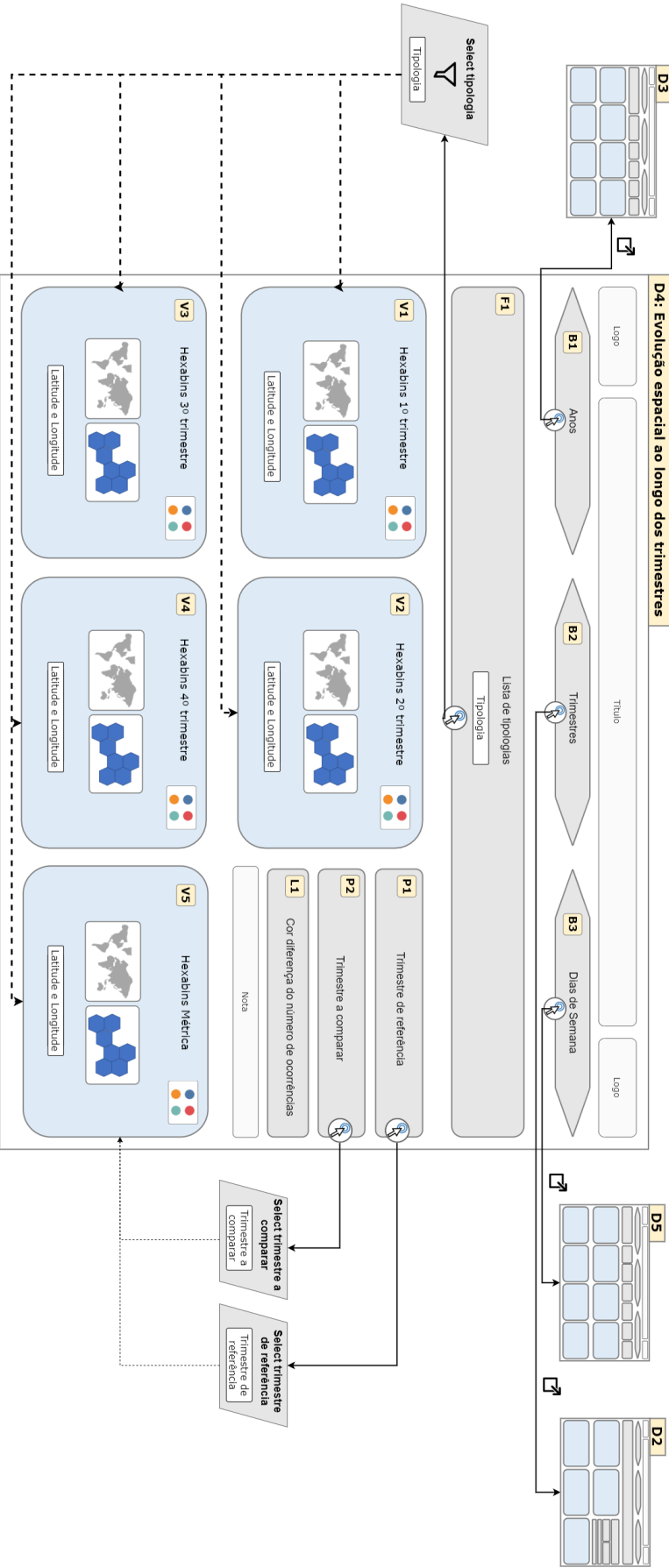


Figura C.7: Diagrama IVML de interações do *dashboard* da evolução espacial das tipologias das ocorrências ao longo dos trimestres.



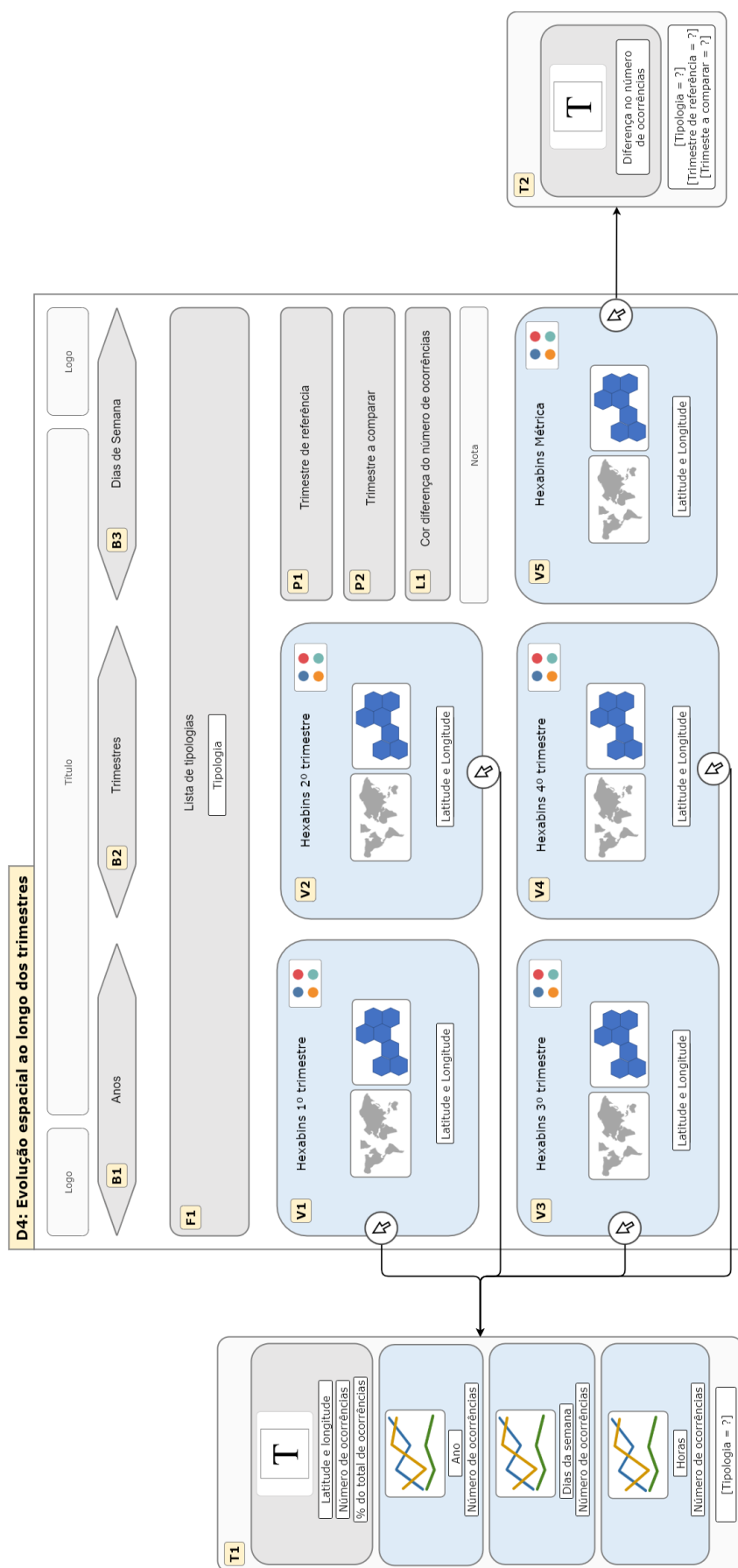


Figura C.8: Diagrama IVML de *tooltips* da evolução espacial das tipologias das ocorrências ao longo dos trimestres.

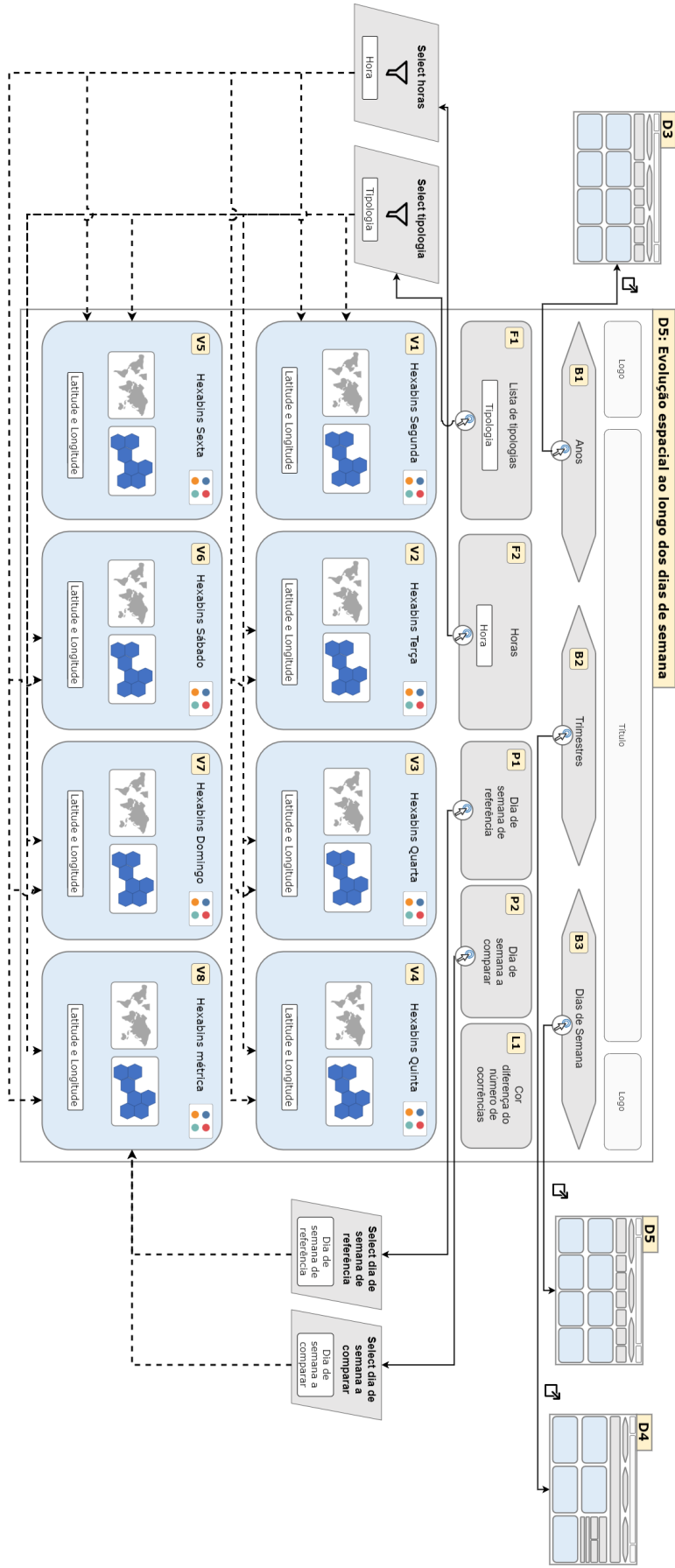


Figura C.9: Diagrama IVML de interações do *dashboard* da evolução espacial das tipologias das ocorrências ao longo dos dias da semana.

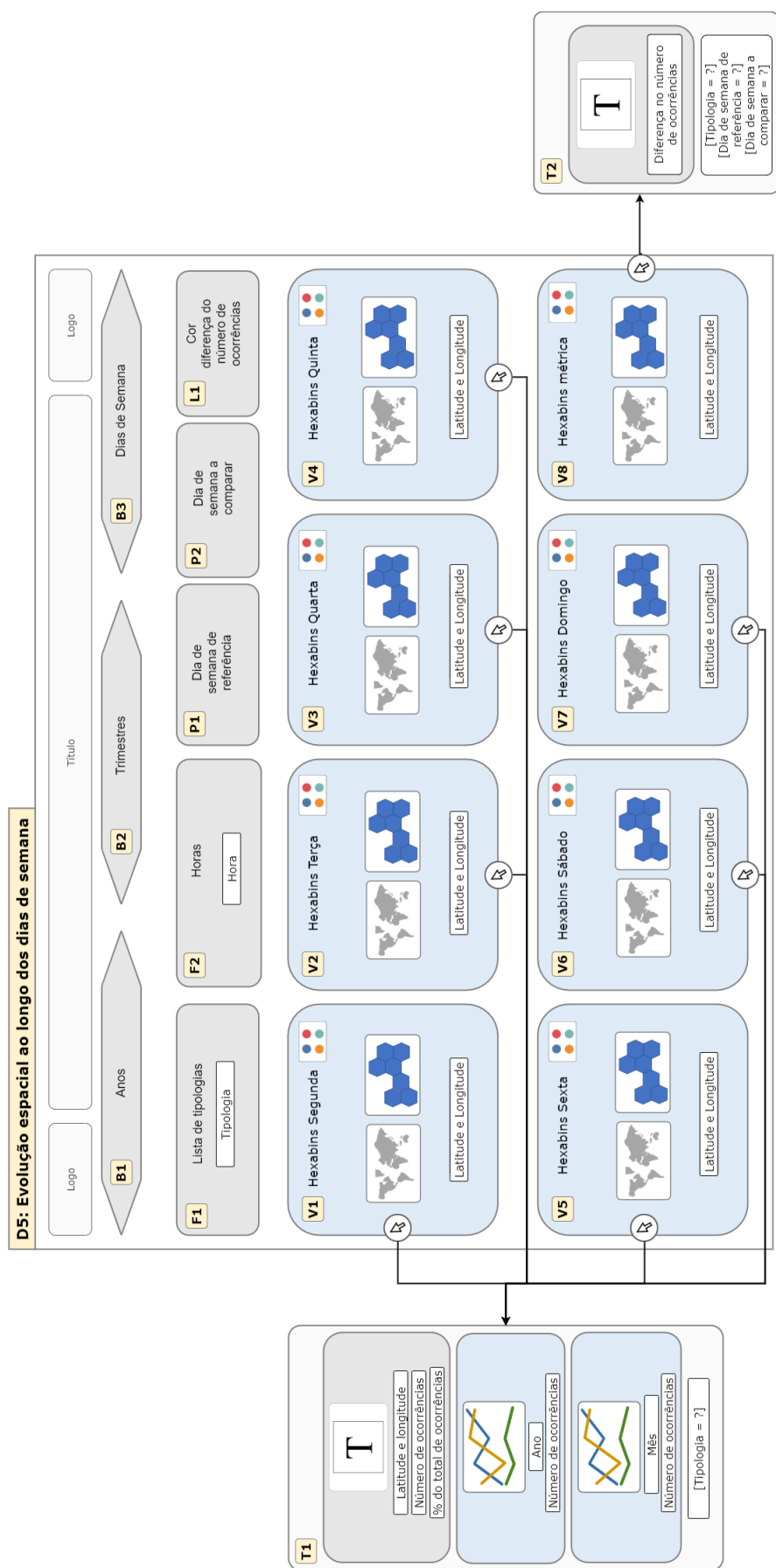


Figura C.10: Diagrama IVML de tooltips do dashboard da evolução espacial das tipologias das ocorrências ao longo dos dias da semana.

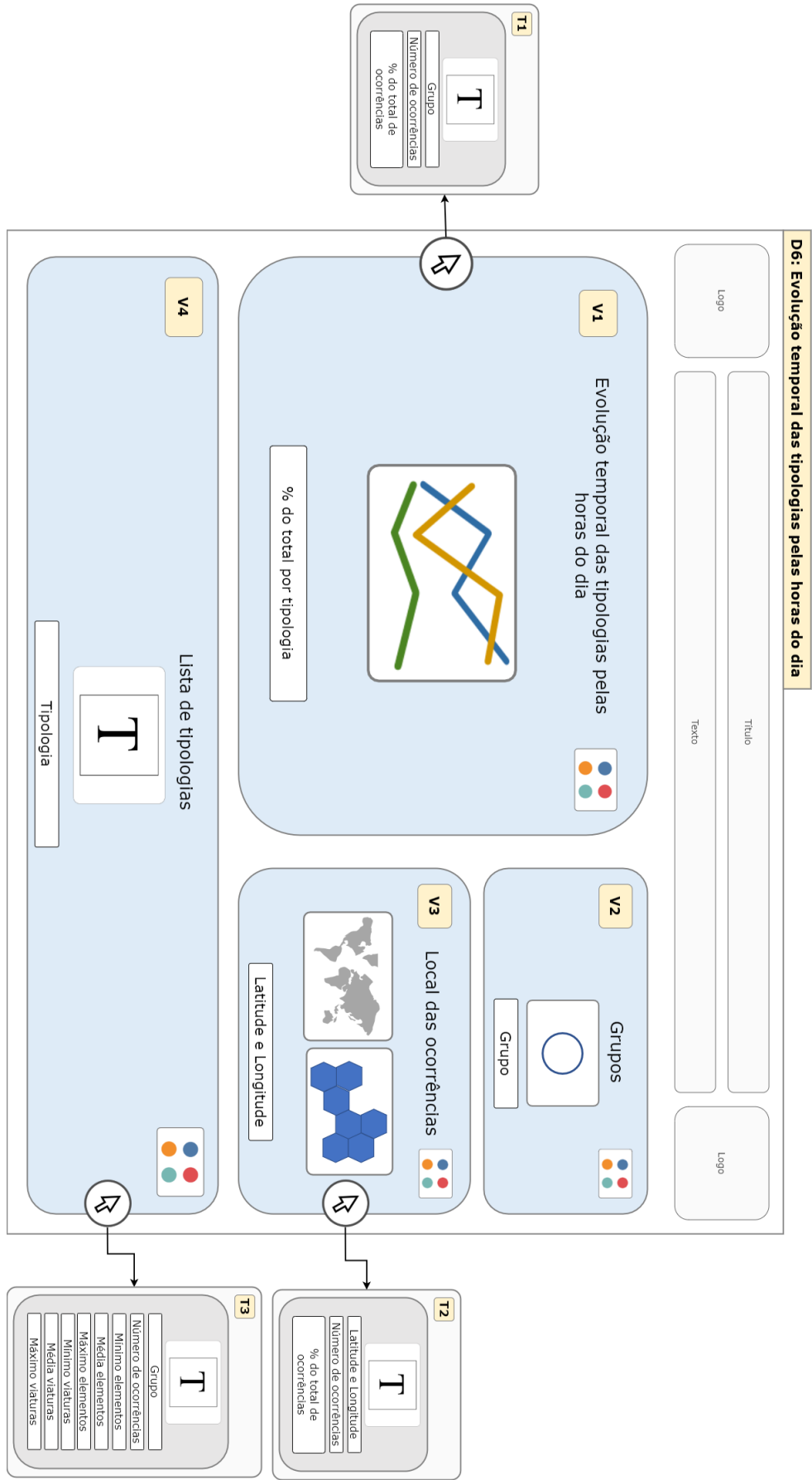


Figura C.11: Diagrama IVML de *tooltips* do *dashboard* de análise do comportamento temporal das tipologias ao longo das estações do ano.

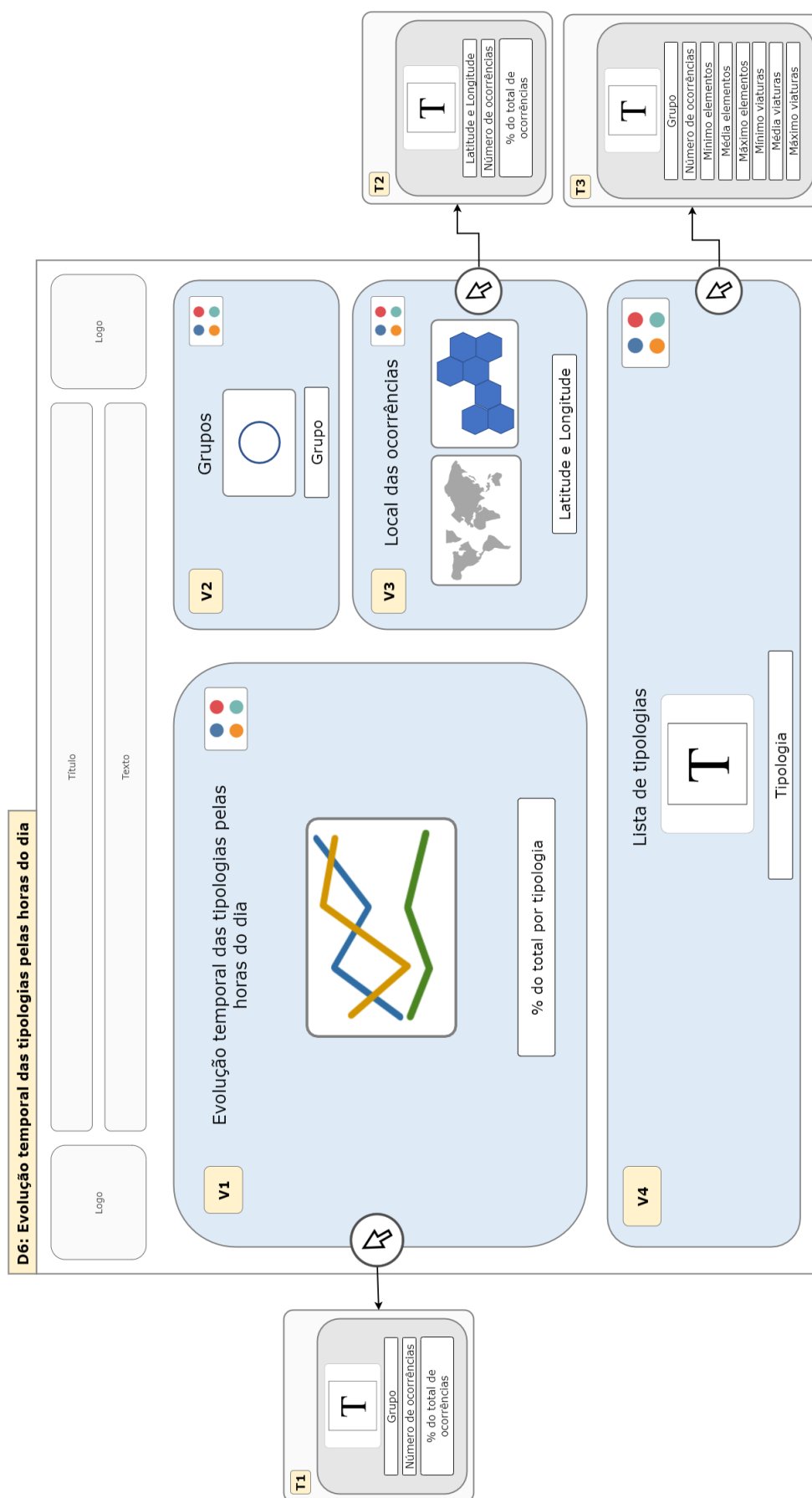


Figura C.12: Diagrama IVML de *tooltips* do *dashboard* de análise do comportamento temporal das tipologias ao longo das estações do ano.

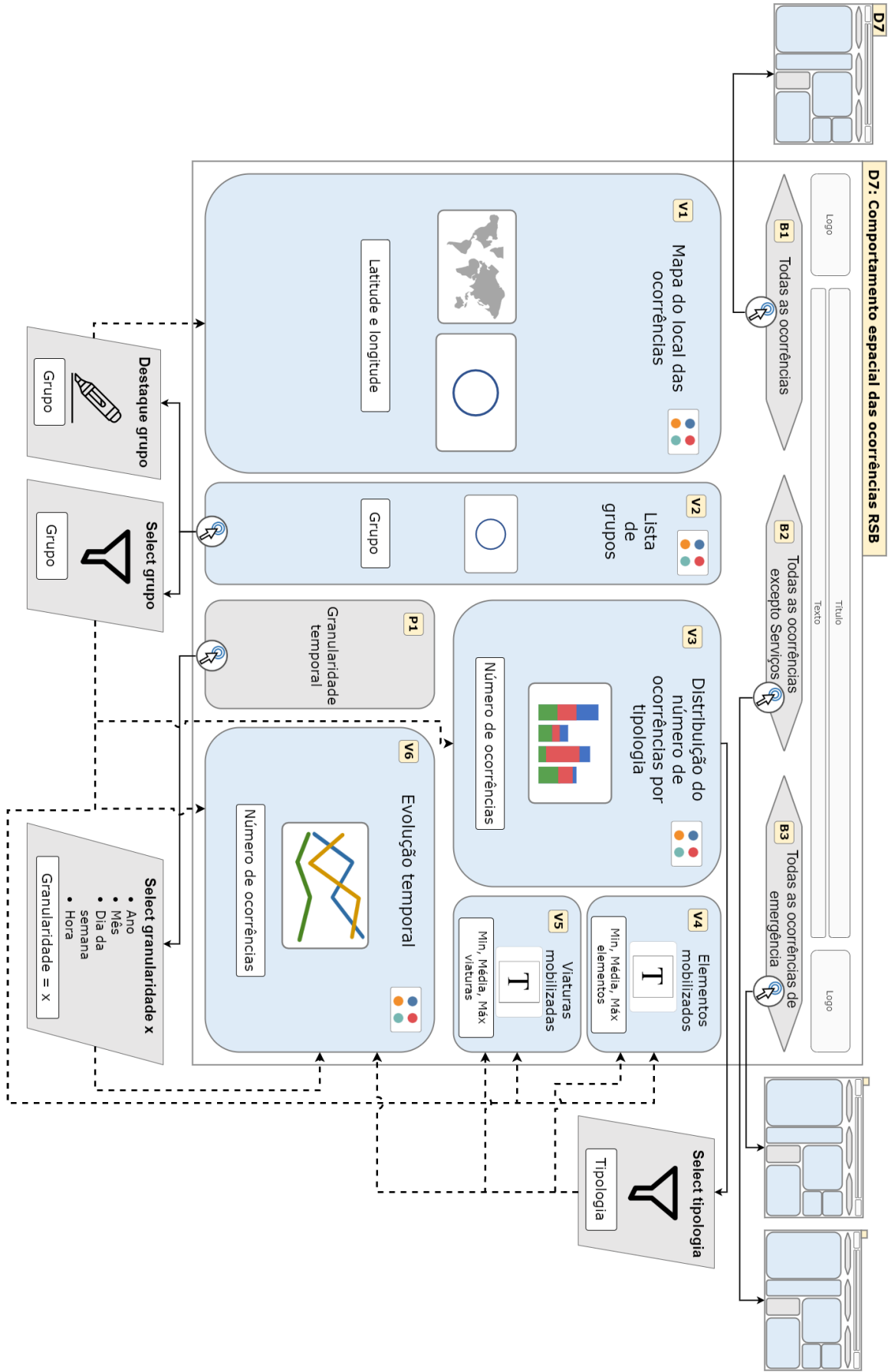


Figura C.13: Diagrama IVML de interações do *dashboard* da evolução espacial das tipologias das ocorrências ao longo do tempo.

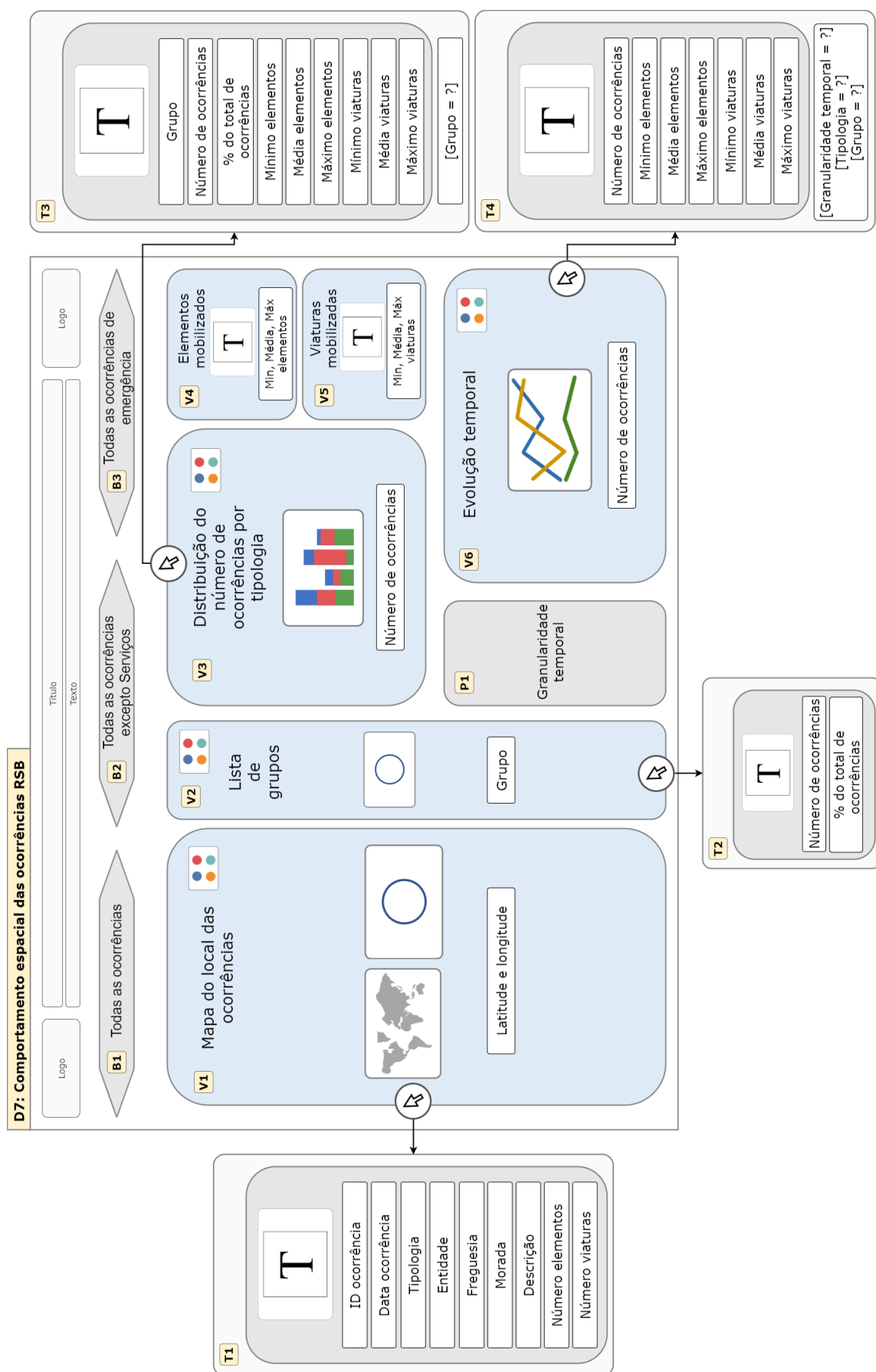


Figura C.14: Diagrama IVML de tooltips da evolução espacial das tipologias das ocorrências ao longo do tempo.

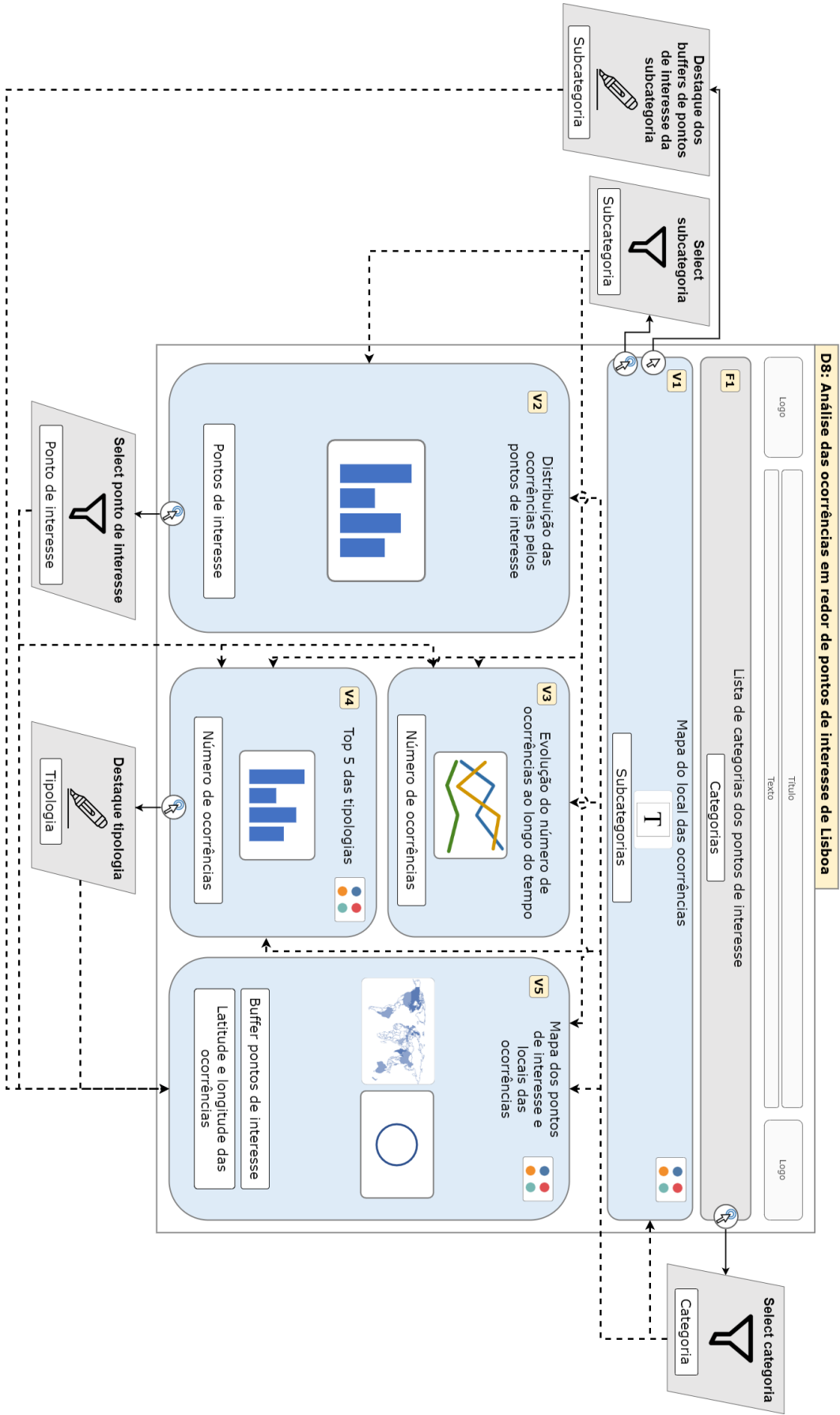


Figura C.15: Diagrama IVML de interações do *dashboard* de análise das ocorrências em redor de pontos de interesse no concelho de Lisboa.



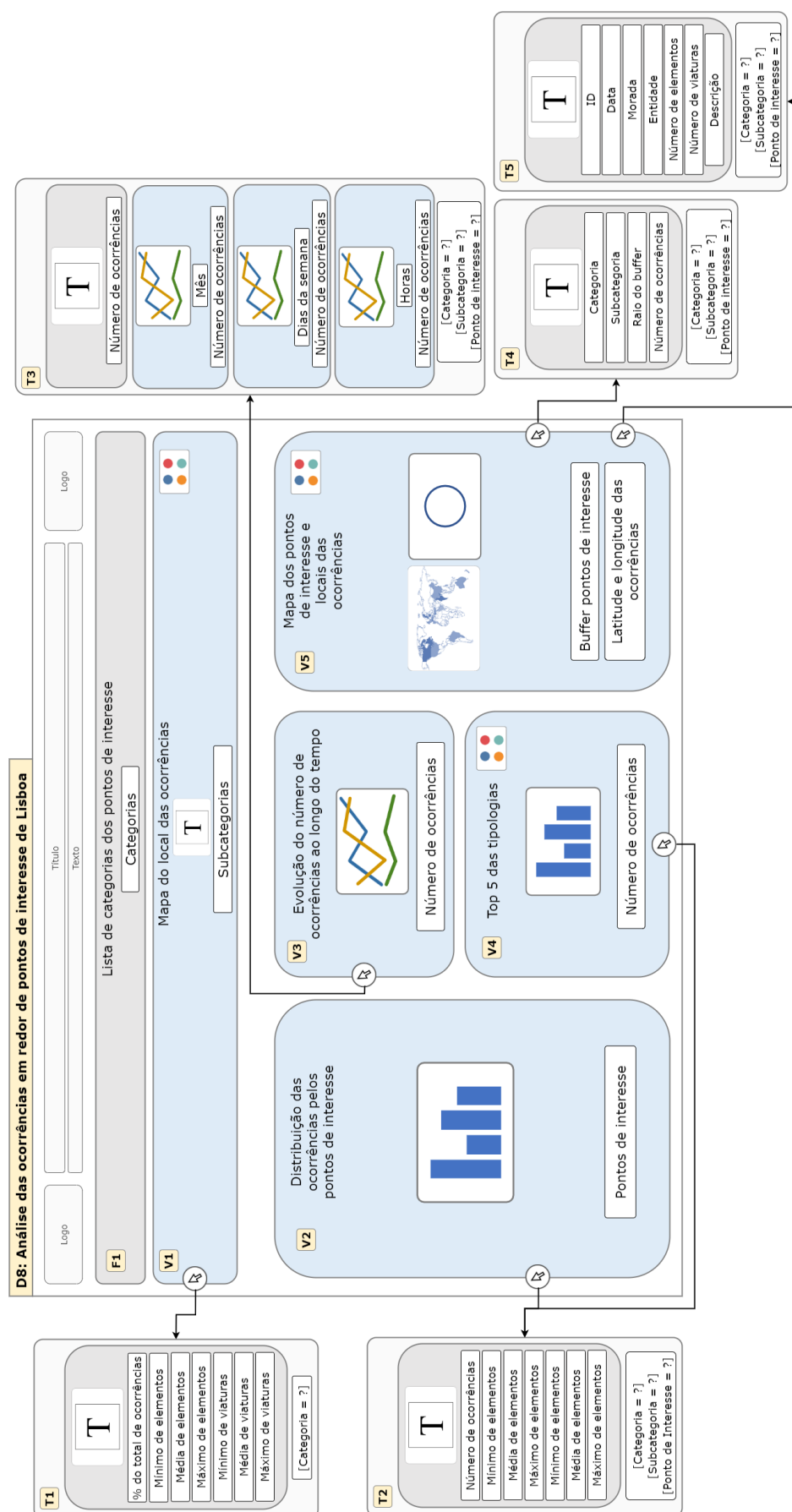


Figura C.16: Diagrama IVML de *tooltips* do *dashboard* de análise das ocorrências em redor de pontos de interesse no concelho de Lisboa.

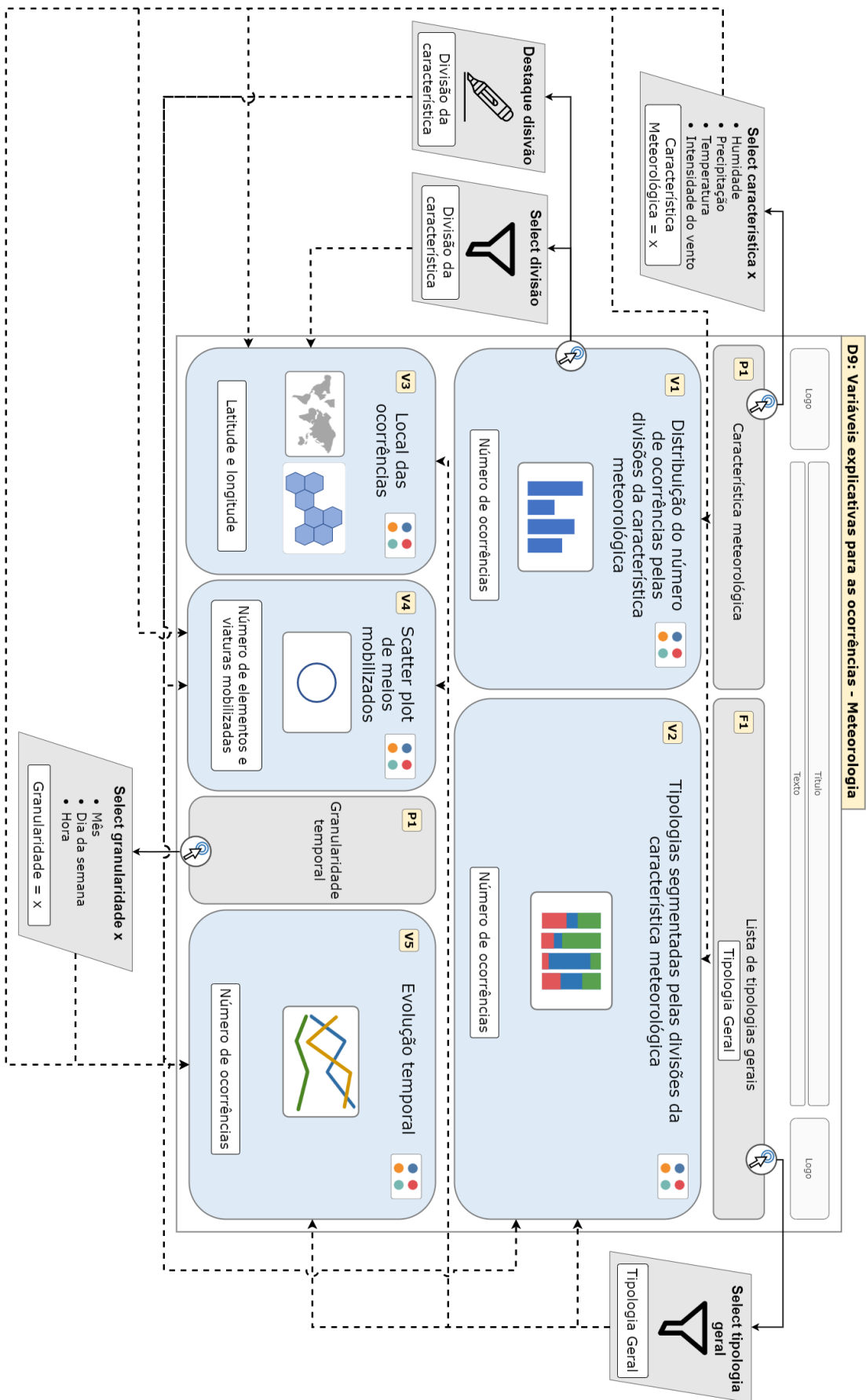


Figura C.17: Diagrama IVML de interações do *dashboard* da procura de variáveis explicativas para as ocorrências do RSB - Meteorologia.

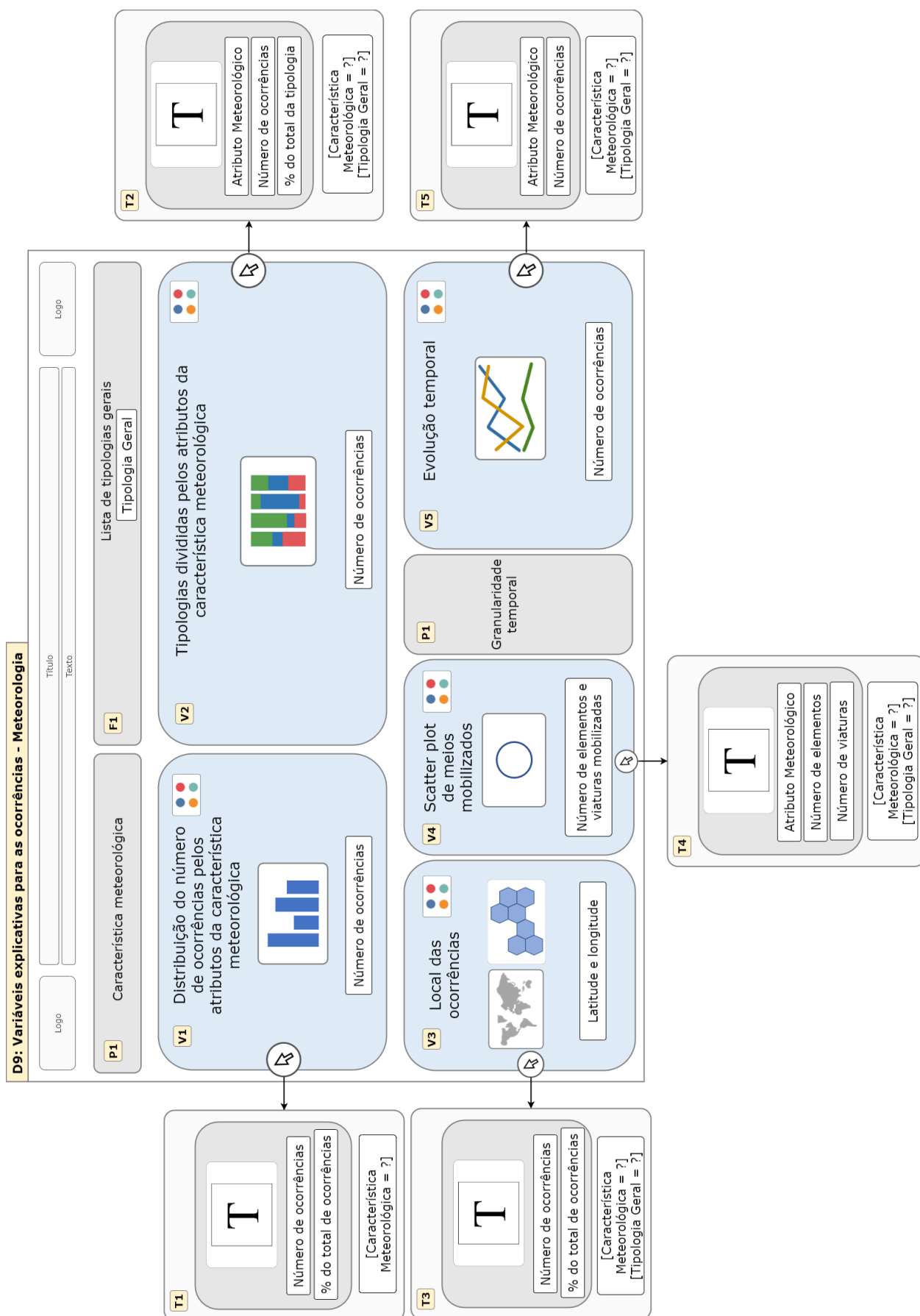


Figura C.18: Diagrama IVML de tooltips do *dashboard* da procura de variáveis explicativas para as ocorrências do RSB - Meteorologia.